THIS WEEK

Room for growth

The European Commission's plans to allow individual countries a veto on the farming of genetically modified crops, although a compromise, should enable the technology to move forward.

When the two camps on either side of a vitriolic debate unite against you, you are probably doing something right — or something horribly wrong. When it comes to acrimonious arguments over genetically modified (GM) crops in Europe, it is hard to be sure, but a move last week by the European Commission does seem to suggest the former.

Last week's political compromise, which should see individual countries able to ban the cultivation of GM crops, even if the crops have been approved at a pan-European Union (EU) level, was attacked by both industry and environmental groups. But some scientists involved in developing and testing the crops were cautiously optimistic that years of rancour have at last yielded to a sensible conclusion.

For years, many European crop researchers have despaired over the hostility to growing GM crops in the region. Although other parts of the world — notably, North America — have sown the seeds and reaped the rewards, the EU has dug itself into an ever deeper hole. Last week's agreement can certainly been seen from two perspectives. National bans that go against the best available evidence about the threat posed by the crops are unfortunate. But, armed with such powers, anti-GM countries should have less incentive to block EU-wide approvals in the first place (see *Nature* http://doi.org/xmq; 2014).

In principle, the EU has a perfectly sensible system for approving new GM crops across the continent. Their safety is assessed by the European Food Standards Agency, which draws up a report for the European Commission. The commission produces a decision that can be discussed by member states, which must then make a final decision by majority. If the member states cannot agree, the final decision is made by the commission. This should take months, not years.

Even those only casually familiar with the EU will see the 'but' coming here. Faced with opposition to GM organisms from member states such as France, and the staunch support of other countries such as the United Kingdom, the commission has sat on approvals, leaving crops and the companies that developed them to languish in a Brussels limbo for years. Companies such as Monsanto have abandoned the EU entirely as far as GM crops are concerned. Research has undoubtedly suffered.

On 3 December, representatives from EU member states and the European Parliament came to a compromise deal. They plan to pass legislation that will allow individual countries to ban crops — something that has been done in the past, but which is a legal grey area. If this agreement clears certain political hurdles, and with nations having the right to stop the use of GM crops in their fields, subject to various provisions, it is to be hoped that the wheels will begin to turn again on the approval process.

Naturally, not everyone is pleased by compromise. Industry groups want a single, uncomplicated market in which to sell their products. Growing and selling GM crops in a fragmented EU will give them a headache. Their opponents in the GM fight are also displeased. The spokesman for the European Parliament's Green grouping said that the agreement could turn into a "Trojan horse", and "could undermine the hand of those wanting to say 'no' to GMOs". The Greenpeace EU Unit said the draft agreement would leave countries that do ban GM organisms open to legal challenges from industry.

Nature has long supported the principle of using GM technology to improve crops (see *Nature* **497,** 5–6; 2013). But it must be acknowledged that a significant proportion of the EU population simply does not want them, for whatever reason. As this journal has also argued, evidence-based policy-making does not always have to side with what the science 'says' is true. It seems correct that countries should have the right to make decisions on this issue that are not based solely on evidence of safety or harm, just as they do on, say, recreational-drug use.

> *"Countries should have the right to make decisions that are not based solely on evidence of safety or harm."*

If the EU's politicians can shepherd last week's agreement into law, at least there will be a way forward. Europe has some highly talented scientists in this field, and they have seen it become increasingly isolated. New technologies are opening up huge opportunities in the genetic engineering of crops, and Europe has already been left behind. But last week's agreement at least shows a willingness to try to catch up. That politicians are willing to compromise on this issue, rather than ignore it, deserves recognition from all sides. ∎

Ethical overkill

Institutions should take a unified look at protections for research on human subjects.

The most important resource needed to conduct research on humans, it is said, is not brainpower or money: it is trust. In the United States, as elsewhere, hundreds of institutions and thousands of investigators work to protect that trust by carefully evaluating proposals for clinical trials and other research that uses human subjects.

Each US institution hosting such a study typically conducts its own ethical review of the proposal. The review process serves many functions: it is an expression of the responsibility that these investigators feel towards protecting their local community, an opportunity to tweak protocols to adapt to the community's specific needs, and a protection against potential lawsuits resulting from a flawed research protocol.

Sadly, evidence suggests that much of this effort is misplaced. A 2010 survey of 45 institutions reviewing the same protocol found that local scrutiny resulted in no substantial changes (B. Ravina *et al. Ann.*

Neurol. **67,** 258–260; 2010). Instead, most alterations simply inserted standardized institutional language — unrelated to the proposed study — to the informed-consent document signed by research participants before they enter a trial. The total cost of all that review: more than US$100,000.

On 3 December, the US National Institutes of Health (NIH) announced a draft policy intended to reduce that redundancy. Open for comment until 29 January, the proposal would require NIH-funded trials that are conducted at more than one site to be approved by a single institutional review board (IRB), which must be willing to shoulder responsibility for all of the sites. The intention is to speed up the approval process for trials that are conducted at multiple facilities. At present, each site may take a crack at reviewing a protocol, often delaying the start of a trial and introducing potential inconsistencies in study protocols and consent forms at different sites.

The NIH's move is the latest in a string of efforts by US regulators to change this institutional practice. In 2006, the US Food and Drug Administration released guidance for clinical trials conducted at multiple sites. In it, the agency stated that this ethical review need not take place at every institution. Instead, each trial could designate an institution to conduct a central review for all participating sites. Four years later, the US Office of Human Research Protections wrote a letter stating its support for that guidance. Despite these assurances, however, it has been difficult to change entrenched institutional practices that have been solidified for more than 40 years.

The NIH's proposal does not prohibit any participating site from conducting its own review, but clearly frowns on the practice — and explicitly pushes the cost of a duplicate review onto the institution.

Inertia is difficult to overcome, particularly at large institutions and with such a valuable resource at stake. Much of this stubbornness is due to an understandable desire by investigators to protect their patients and community. Some local IRB members feel that abdicating their review of research protocols is a violation of their responsibility to that community, and worry that standards will slip if they do not personally review the study.

"There is no evidence that multiple ethics reviews enhance protections for human subjects."

As the NIH has said, there is no evidence that multiple ethics reviews enhance protections for human subjects. Centralized review may seem to save time and money, but there is no clear evidence that it protects study subjects any better. Still, the NIH's move to encourage central review is the right one, given the available evidence.

Regulations that favoured local IRB reviews were developed in an era when studies were typically done at a single site. This is no longer the case. As therapies become more tailored to individual genetics, and diseases are subdivided into rarer subtypes, more sites are needed to enrol enough patients to evaluate an intervention.

Around the world, DNA sequencing labs are generating reams of genetic data that could hold the clues to the next medical revolution. Finding those clues quickly and ethically will require studies that combine data from across the globe. Investigators are clamouring for unified informed-consent documents that will allow them to compile genetic information into databases without creating a legal thicket of differing privacy protections. The NIH's move is an important step in that direction, but there is much farther to go. ∎

Protect and serve

Nations must keep expanding conservation efforts to avoid a biodiversity crisis.

There are 22,413 species deemed at risk of extinction by the International Union for Conservation of Nature (IUCN). If some ambitious person tried to read out their names — without any breaks for food or water — it would take at least half a day. But that would be just the start. The IUCN has assessed the status of only 76,199 of the 1.7 million species of animals, plants, fungi and protists on Earth that have been described by scientists. And some suggest that at least five times more species still wait to be discovered. Many of these are also threatened, and it would take months to read out all of their names. (Except that they do not, of course, have names.)

There remain vast gaps in knowledge about the planet's biodiversity — and the precarious state of life. Every day, animals and plants go extinct. Nobody knows exactly how many, but estimates range from 500 to 36,000 extinctions per year. A News Feature on page 158 draws together some of the best studies of biodiversity and tries to make such vast numbers fathomable.

Before human populations swelled to the point at which we could denude whole forests and wipe out entire animal populations, extinction rates were at least ten times lower. And the future does not look any brighter. Climate change and the spread of invasive species (often facilitated by humans) will drive extinction rates only higher.

The pace of extinction is leading towards a crisis. If all currently threatened species were to go extinct in a few centuries and that rate continued, the die-offs would soon reach the level of a mass extinction — the kind of biological catastrophe that ended the reign of the dinosaurs and that has happened only five times in Earth's history. The sixth mass extinction could come in a couple of centuries or a few millennia, but it lies somewhere in the future if nations keep to their present course.

There are some hopeful signs. Countries are rapidly expanding the areas they shield from destructive human activities. The United Nations Environment Programme (UNEP) announced last month that countries have set aside 6.1 million square kilometres of ocean and land habitat since 2010, which increases the total protected areas to 15.4% of Earth's land and 3.4% of its oceans. According to UNEP, countries are on track to meet a 2020 goal established under the Convention on Biological Diversity to protect 17% of land areas, although reaching the 10% target for coastal and marine regions will require further efforts. The total areas set aside now equal the size of Africa.

But these efforts are not enough. Many protected zones are 'paper parks', where hunting, fishing and habitat destruction continue apace because of lax enforcement. And most parks established so far do not protect the most crucial areas — the ones full of threatened species and habitats. Nations are also investing much less on protection than they were 15 years ago, after adjustments are made for inflation.

In the face of this uncertainty about biodiversity, what should the world do? UNEP estimates that it would take US$76 billion each year to establish and manage a set of expanded parks that protect important habitats for all wildlife groups. That figure is just as unfathomable as the number of species on the planet. But consider that a blockbuster video game can sell $500 million in copies in a single day. According to UNEP, the economic benefits of protected areas far outweigh their costs, which could be met through a mixture of conventional sources and innovative funding mechanisms, such as green taxes and payments for the services that ecosystems provide.

As part of this protection effort, nations also need to devote more resources to taking stock of life. The IUCN has set a 2020 goal of assessing 160,000 species, roughly double the current number, which it calculates would cost $60 million and cover a good representation of most major taxonomic groups and ecosystems. The job of counting and evaluating is not the most exciting science. But it is one of the most fundamental and important tasks that humans can do — to take a measure of life and protect what remains before it disappears. ∎

↪ NATURE.COM
To comment online, click on Editorials at: **go.nature.com/xhunqv**

Assess the real cost of research assessment

The Research Excellence Framework keeps UK science sharp, but the process is overly burdensome for institutions, says **Peter M. Atkinson**.

'Twas the week before Christmas, and all through the United Kingdom, scientists were waiting nervously to see how many glittering prizes the government would stuff into their stockings. Those prizes — the results of the Research Excellence Framework (REF) exercise, to be announced on 18 December — will go some way towards determining which researchers in UK universities have a happy New Year.

The scale and importance of this assessment of publicly funded research is unique to the United Kingdom. Run every five years or so, the REF system grades the quality of research in dozens of fields across more than 100 institutions, and allocates government grant money accordingly. The winners enjoy high-quality ratings for their academic departments and the guarantee of a hefty chunk of cash to support their research. A poor rating can see a department starved of money or even closed down.

The government argues that this regular scrutiny has helped to consolidate the United Kingdom's place as a global scientific superpower. And an institution with an excellent rating in physics, say, or chemistry can use it to attract staff and students. But the REF comes at a heavy cost — the amount of time and work it takes institutions and staff to prepare submissions.

Work is already under way to prepare for the next exercise, expected to run in 2020. All involved should also start to think about how to do it differently, to keep the good points but minimize the workload.

Perhaps the largest burden for institutions is that of choosing which researchers will represent each subject in the assessment. Although it is departments and disciplines that are ultimately graded, their grades are based mainly on the outputs of individuals who work in them. But there is a tension here. Funding is per head, so of two equally rated departments, the one that submits the work of more researchers receives more money. But as the number of scientists included goes up, the overall quality of the research submitted goes down — even the very best departments have a limited number of truly world-leading researchers.

A chemistry department of 60 researchers, for example, can agonize over whether to submit the research of 50 or 40 of them. To make the decision, it will do its own assessment of the quality of each scientist's work, then rank the results and try to calculate where to draw the line between who is submitted and who is not. The department must not only grade the research of its own scientists, but also grade it according to how it thinks the REF will do so. The department must also consider where

departments at rival institutions are likely to draw their own lines. But, of course, there is rarely any information on a competitor's strategy. So game theory comes into play, but with few data to drive decision-making.

In my own research, I have found that such judgements are imprecise and vary to a large degree. Why? Because uncertainty is always present. Researchers asked to rate the quality of a colleague's work, from 0 to 10, for example, will rarely come up with the same score, and this uncertainty makes internal selection all the harder. Where does this leave the REF? Although the overall strategic effect of the exercise has been positive for the quality of UK science, the amount of effort it requires of institutions deserves a rethink.

More of the process could be automated, using 'big data' and bibliometric and machine-learning approaches. To reduce the workload on institutions — most of which already subscribe to systems that capture the real-time information needed — the REF should assess the outputs of all eligible staff, removing much of the selection burden. A machine cannot yet judge the quality of research output, but there are surrogates. For many subjects, bibliometric analysis can leverage the peer-review process that already occurs through publication, as well as the peer assessment implicit in citation data. (An independent review of the use of such metrics in a future REF was launched this year.)

The REF includes other subjective judgements of quality, including — for the first time this year — the socio-economic impact of research. These impact reports are written specifically for the REF and so add considerable effort to the process. And it is arguably harder for the REF to judge and compare quality in this area. There is no guarantee, for example, that a spin-off company that generates 200 jobs and £20 million (US\$31 million) in investments will be judged to have more impact than a spin-off that generates 20 jobs and £2 million in investments. Automation is not possible here, but there is room for greater standardization of the dimensions by which impact is assessed and the criteria against which quality is judged.

As institutional access to big data increases and technology improves, it makes sense to use all the data available to inform judgements. An obvious benefit is that the REF could be updated annually on the basis of an electronic snapshot. These changes would not make the REF perfect, but it is not perfect now. They would, however, reduce its burden and allow institutions to focus on research. ∎

> ## THE AMOUNT OF EFFORT THAT THE REF REQUIRES OF INSTITUTIONS DESERVES A RETHINK.

Peter M. Atkinson *is professor of geography at the University of Southampton, UK.*
e-mail: p.m.atkinson@soton.ac.uk

RESEARCH HIGHLIGHTS
Selections from the scientific literature

Smartphones sniff gases

A common technology that enables short-range communication in smartphones could be used to detect airborne chemicals.

Near-field communication chips are found in half a billion mobile devices worldwide. They communicate wirelessly with small external tags and are used in contactless payment systems, for instance. A team at the Massachusetts Institute of Technology in Cambridge, led by Timothy Swager, modified the circuitry in the external tags using nanomaterials that are sensitive to certain chemicals. When a particular gas is present, the tag short-circuits and the smartphone can no longer read the tag.

By scanning combinations of tags, each of which was sensitive to a different chemical, the team could distinguish between gases including ammonia, hydrogen-peroxide vapour and water vapour — down to the level of parts per million.

Such a system could be used to detect explosives or pollution and has other applications, the authors say.
Proc. Natl Acad. Sci. USA **http://dx.doi.org/10.1073/pnas.1415403111 (2014)**

Ancient apes digested ethanol

Human ancestors were able to metabolize ethanol 10 million years ago, around the time that they came down from the trees.

Matthew Carrigan at Santa Fe College in Gainesville, Florida, and his co-workers analysed the gene encoding the enzyme ADH4, which is made in the digestive tract to metabolize ethanol. They studied this gene from 28 mammals, including 17 primates, to trace its 70-million-year evolutionary history.

When they synthesized various ancestral forms of the enzyme, they found that ADH4 from ancestors of humans, chimpanzees and gorillas broke down ethanol much more efficiently than the enzyme from more ancient ancestors.

This change might have helped the hominids adapt to life on the forest floor, where there was probably more fermented fruit than in trees.
Proc. Natl Acad. Sci. USA **http://doi.org/xkp (2014)**

Cockroach night-vision

Cockroaches can see in near-darkness thanks to the many light-sensing cells in their eyes that pool a tiny number of light signals over space and time.

Matti Weckström and his colleagues at the University of Oulu, Finland, tested the behavioural responses of the American cockroach (*Periplaneta americana*) to varying levels of light, using a virtual-reality system that displayed moving patterns (pictured). By recording from individual light-sensitive eye cells, they found that each photoreceptor receives only one photon every 10 seconds when light levels are equivalent to a moonless night, during which the animals could still see. This pooling probably occurs over thousands of photoreceptors in the eye, say the authors.

Further study might improve night-vision devices, they add.
J. Exp. Biol. **217, 4262–4268 (2014)**

Antarctic ice loss accelerates

Glaciers flowing into Antarctica's Amundsen Sea are some of the fastest melting on the continent — and in recent years have lost ice at an ever-quicker rate.

Different remote-sensing techniques have yielded slightly different estimates for the amount of ice melting from the Amundsen glaciers. Tyler Sutterley of the University of California, Irvine, and his colleagues compared and reconciled four ice-measuring methods. They found that between 2003 and 2009, the disappearance of Amundsen ice accelerated at a rate nearly three times faster than over the whole period between 1992 and 2013.

The findings boost confidence in the various ice-measuring methods and confirm just how quickly these glaciers are funnelling ice into the sea.
Geophys. Res. Lett. **http://doi.org/xms (2014)**

Melted Antarctic ice may not return

The melting of ice around Antarctica as a result of global warming could be irreversible.

Jeff Ridley and Helene Hewitt of the UK Met Office's Hadley Centre in Exeter used a global climate model to examine how polar sea ice responds to changing climates. They found that Arctic sea ice melts and reforms in response

to changing temperatures when carbon dioxide concentrations in the models are first increased and then gradually reduced to pre-industrial levels. In Antarctica, however, sea ice returns at first, but had not recovered by the end of the simulation, even after a further 150 years of pre-industrial CO_2 levels.

This lack of ice recovery is a result of strong heat uptake by the Southern Ocean, which continues to warm parts of the seas around Antarctica long after global warming has been reversed, according to the authors.

Geophys. Res. Lett. http://doi.org/xh3 (2014)

ANIMAL BEHAVIOUR

Electric eel zaps neurons of its prey

The electric eel stuns its fish prey by emitting electrical pulses that control parts of the nervous system of its victim.

Kenneth Catania at Vanderbilt University in Nashville, Tennessee, studied the behaviour and electrical discharges of an eel (*Electrophorus electricus*; **pictured**) when it was presented with fish in an aquarium. He found that the eel's shocks immobilize the fish by activating nerves controlling the muscles, causing them to contract throughout the fish's body even when the fish's brain and spinal cord were destroyed. When the fish was hidden, the eel sent out two quick pulses,

causing the fish to twitch, followed soon by a high-voltage zap and an attack.

The results show how the electric eel can remotely control its prey.
Science 346, 1231–1234 (2014)

IMMUNOLOGY

How immune cells search and destroy

To locate the source of an infection, immune cells called neutrophils take directions from local blood cells.

Neutrophils are the first responders to an infection, where they produce pathogen-killing compounds. To determine how they home in on infections and other injuries, a team led by Andrés Hidalgo at Spain's National Centre for Cardiovascular Research in Madrid imaged blood vessels in live mice that were showing an inflammatory response.

The authors discovered that neutrophils drifting in the bloodstream stuck to blood vessel walls and then sent out arm-like extensions. When these encountered blood cells called platelets — which are activated by injury to help to stop bleeding — the neutrophils began to migrate along the vessel wall and churn out toxic chemicals. Blocking communication between neutrophils and platelets lessened tissue damage from excessive inflammation in mouse models of sepsis, lung injury or stroke.
Science 346, 1234–1238 (2014)

NEUROSCIENCE

Injury blunts brain waste disposal

Fluid channels in the brain that help to remove waste could be impaired after traumatic injury, promoting cell death.

After injury, brain cells release a protein called tau, which accumulates as tangles and is associated with neurodegeneration and dementia. Jeffrey Iliff at the Oregon Health and Science

University, Portland, and his colleagues showed that tau is cleared from young healthy mouse brains along the 'glymphatic pathway', channels that wash out waste from the brain.

The authors found that after traumatic injury, the pathway's performance decreased by about 60%. It was reduced even further in injured mice in which a gene important for the pathway, aquaporin-4, had been knocked out. These mice developed tau tangles and performed less well in cognitive tests.
J. Neurosci. 34, 16180–16193 (2014)

ANIMAL BEHAVIOUR

Some bats click wings to navigate

Some bat species unable to use sonar to sense their environment can instead navigate using echoes from clicking their wings — possibly an early, crude form of echolocation.

A team led by Arjan

Boonman and Yossi Yovel at Tel Aviv University in Israel studied three species of wild, non-echolocating Old World fruit bat (**pictured** is *Cynopterus brachyotis*). They found that individuals of two species emitted clicks more frequently in the dark than in the light, and could find and land on large objects, although they failed to detect small obstacles. When the researchers taped the bats' wings, the clicking stopped, but the exact clicking mechanism could not be determined.

The authors suggest that much can be learned about the evolution of echolocation from these fruit bats.
Curr. Biol. http://doi.org/xmr (2014)

↪ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

FACILITIES

Megascope member

India announced on 2 December that it will become a full partner in the Thirty Meter Telescope, joining a consortium that includes institutions from China, Japan and the United States. The deal secures Indian scientists time on the next-generation telescope, which will be one of the world's largest when it opens on Mauna Kea in Hawaii — scheduled for the 2020s. Last week, another organization gave the green light to construction of the 39-metre European Extremely Large Telescope on Cerro Armazones in Chile.

Cell institute

On 8 December, billionaire philanthropist Paul Allen announced plans to invest US$100 million to create the Allen Institute for Cell Science, modelled on the Allen Institute for Brain Science in Seattle, Washington. Cell biologist Rick Horwitz will lead the institute, which will also be located in Seattle. The centre will develop a 'cell observatory' to display how a cell's components work together. See page 157 for more.

Rocket ramp-up

Europe will press ahead with developing a cheaper type of rocket for satellite launches, thanks to a funding agreement reached by the 20 member states of the European Space Agency (ESA) on 2 December. The Ariane 6 will replace the Ariane 5, which faces increasing industry competition from rockets built by start-up companies such as SpaceX of Hawthorne, California. ESA will spend an estimated €3.8 billion (US$4.7 billion) on the new designs, which include upgrading the smaller Vega C rocket.

NASA's Orion test flight soars

NASA's next-generation vehicle for sending astronauts to deep space made its inaugural flight on 5 December in a spectacular morning launch (pictured) from Cape Canaveral, Florida. In an uncrewed test to see how its systems would fare in high-radiation environments, the Orion capsule made nearly two full orbits of Earth before splashing down in the eastern Pacific Ocean (see go.nature. com/zmwarj). At its highest, Orion flew 5,800 kilometres from Earth, the farthest that any human-rated space vehicle has been since the final US lunar-landing mission, *Apollo 17*, in 1972. In other launch news, on 3 December the Japan Aerospace Exploration Agency succeeded in sending its Hayabusa-2 probe off on a journey to collect samples from an asteroid and return them to Earth.

EVENTS

Einstein's reams

Thousands of Albert Einstein's letters and writings are freely available online through a website launched 5 December. The site is a partnership involving Princeton University Press in New Jersey, the Rhode Island digital publisher Tizra, the Hebrew University of Jerusalem and the California Institute of Technology in Pasadena. The Digital Einstein Papers (go.nature.com/grg6rh) contain 5,000 documents transcribed and translated to English that span the first 44 years of Einstein's life. Eventually, the repository will include all of the physicist's archived papers.

Nobel sale

On 4 December, a buyer paid US$4.1 million for James Watson's Nobel prize medallion. Media reports say that the buyer, Russian billionaire Alisher Usmanov, plans to return the medal to Watson, who shared the 1962 Nobel Prize in Physiology or Medicine for co-discovering the double-helix structure of DNA and is the first scientist to auction his own Nobel medal. In 2007, he retired from Cold Spring Harbor Laboratory in New York, after generating friction with his suggestions that black people are not as intelligent as white people. See go.nature.com/t4ejud for more.

POLICY

GM crop bans

European Union (EU) politicians have reached an agreement that, if passed into law, could allow the cultivation of new genetically

modified (GM) crops in the EU. Representatives of member states and the European Parliament decided on 3 December to allow individual nations to ban GM crops for cultivation, even if they have been approved by the EU. Approvals had stalled for years as pro-GM governments such as that of the United Kingdom clashed with anti-GM nations such as France. See go.nature.com/5lzdsn for more.

Chimp, not human

A New York appeals court has refused to grant legal personhood to Tommy, a captive chimpanzee. A group called the Nonhuman Rights Project has fought to free Tommy and other chimps, including two research animals, by arguing that the chimps deserve the human right of bodily freedom. Lower courts rejected the Florida-based organization's lawsuits last year; the first appeal was shot down on 4 December. The organization is pushing ahead with other appeals, and says that it will take Tommy's case to New York's highest court.

Critical habitat

The US National Oceanic and Atmospheric Administration (NOAA) Fisheries on 2 December proposed to designate a critical habitat of more than 906,000 square kilometres of the Bering,

Chukchi and Beaufort seas for the Arctic ringed seal (*Phoca hispida hispida*). Shrinking sea ice and declining snowfall are threatening the animals (**pictured**), which nurture their pups in snow caves and use ice platforms for moulting and other activities. Under the proposed status, federal agencies that fund or authorize activities in the habitat (such as oil drilling) would have to consult NOAA Fisheries first.

UK science budget

The UK Chancellor of the Exchequer George Osborne stressed the importance of science last week. In his autumn budget statement on future government spending, he introduced student loans to fund master's degrees and measures to increase tax credits for companies investing in research. Osborne also warned that there would be more cuts to public spending if his Conservative party maintained power after a 2015 election, suggesting that the core UK science budget could continue to fall in real terms.

Trials streamlined

Conducting clinical trials at multiple US sites may become easier under a draft policy released by the US National Institutes of Health (NIH) on 3 December. Currently, studies that use human participants must meet the ethical, safety and informed-consent requirements of the institutional review board at each site — but the rules can vary widely. The NIH proposal would allow a single board to oversee all centres involved in a trial, which the agency says would reduce paperwork and expedite research. The draft policy is open for comments until 29 January.

Energy leader

On 8 December, the US Senate confirmed physicist Ellen Williams as director of the Department of Energy's Advanced Research Projects

15–19 DECEMBER
Scientists meet in San Francisco, California, to discuss the latest research in Earth, ocean and planetary sciences at the American Geophysical Union's Fall Meeting.
go.nature.com/ylqer9

Agency — Energy (ARPA-E), where she will oversee a new programme to fund promising energy technologies that are still too young for private-sector investment. Williams is currently on leave from the University of Maryland in College Park, and became the chief scientist for British oil-and-gas company BP in 2010.

Antibody advance

On 3 December, US regulators approved blinatumomab, the first of a new generation of therapeutic antibodies that bind to multiple targets. The cancer-fighting drug, made by Amgen of Thousand Oaks, California, will be marketed for treating a rare form of acute leukaemia. It works by tethering immune cells called T cells to cancer cells, triggering the T cell to attack.

Antibiotics deal

Pharmaceutical giant Merck of Whitehouse Station, New Jersey, is going into the antibiotics business. On 8 December, the company announced that it was acquiring Cubist Pharmaceuticals, based in Lexington, Massachusetts, for US$8.4 billion. Cubist, which specializes in antibiotics to treat drug-resistant infections, has received fast-track approval from the US Food and Drug Administration for several drugs currently under development.

TREND WATCH

The global average temperature is headed for a record high this year, according to measurements averaged from the UK Met Office and the University of East Anglia's Climatic Research Unit, the US National Oceanic and Atmospheric Administration's National Climatic Data Center and NASA's Goddard Institute for Space Studies (see chart). In a 3 December report, the World Meteorological Organization in Switzerland highlighted severe flooding in 2014 in South Africa, northern Pakistan and India.

RISING HEAT
Averaged data from January to October show that 2014 is on track to be the warmest year, or one of the warmest, on record.
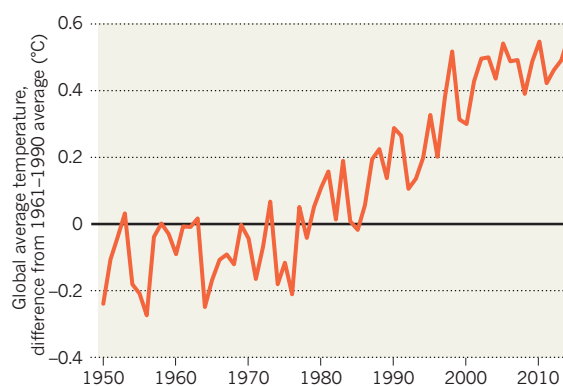
NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

An elephant pulls debris near the coast of Banda Aceh in Indonesia, after the 2004 Boxing Day tsunami.

EARTHQUAKES

Tsunami alerts fall short

Ten years after the devastating Sumatra earthquake, warnings for the Indian Ocean go out, but often fail to reach the people most at risk.

BY ALEXANDRA WITZE

When a magnitude-9.1 earthquake shuddered to life off the Sumatran coast on 26 December 2004, there was no systematic way to alert communities across the Indian Ocean that a devastating wave might be coming. Afterwards, with some 230,000 people dead and US$14 billion in damages, international disaster experts resolved to reduce the toll next time a tsunami struck.

Ten years on from the deadliest tsunami in history, almost all the countries bordering the Indian Ocean are hooked into a network of seismometers, sea-level gauges and satellite-linked buoys. In close to real time, this Indian Ocean Tsunami Warning and Mitigation System (IOTWS) notifies nations from Indonesia to Sri Lanka to Oman when a big offshore earthquake has occurred and determines whether it might generate a tsunami. Were the 2004 earthquake to happen today,

these nations would be much better prepared.

But despite its technical sophistication, the tsunami warning system remains vulnerable. The initial rush of funding from international donors is drying up, and Indian Ocean nations now face the responsibility of maintaining the system — to the tune of between $50 million and $100 million per year. "We're definitely safer than we were in 2004," says Rick Bailey, head of tsunami warning services at the Australian Bureau of Meteorology in ▶

EARLY WARNING

A network of seismometers, coastal sea-level gauges and offshore tsunameters has been established in the decade since the 2004 Indian Ocean tsunami that killed more than 200,000 people. The network can issue notices of an approaching tsunami that give people ample time to evacuate coastal areas, but getting those warnings to remote locations has been a challenge.

2004 tsunami epicentre

Indian Ocean

- Broadband seismometer
- Sea-level gauge
- Tsunameter
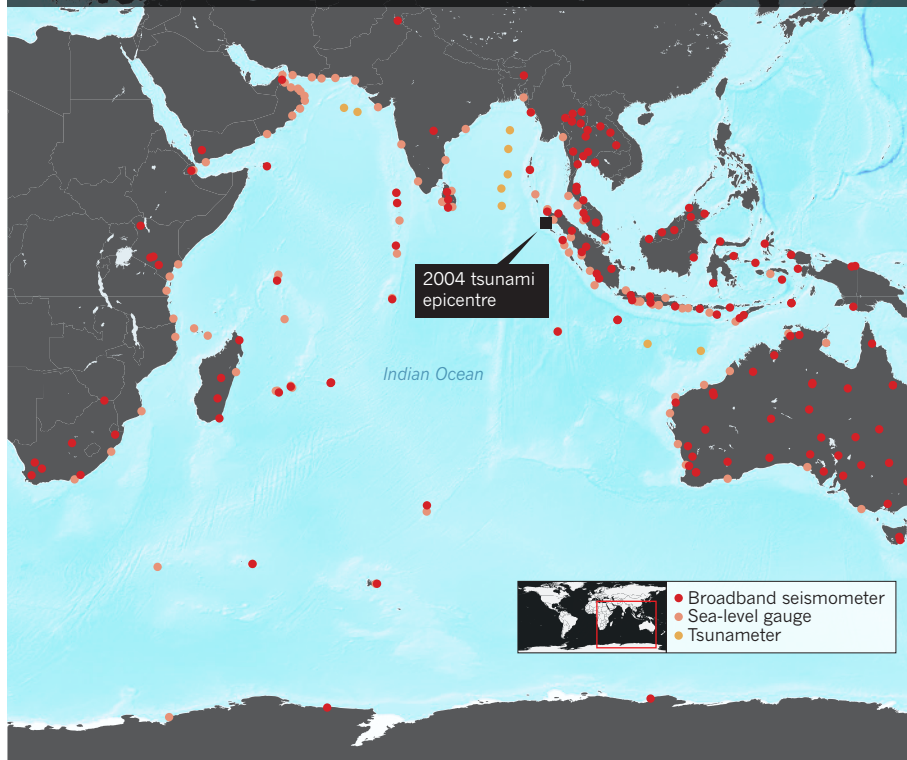
▶ Melbourne. "But sustainability will be the next big issue for us."

The geophysical components of the Indian Ocean tsunami-alert system generally work well (see 'Early warning'). More than 140 seismometers constantly monitor earthquakes around the basin, including in the quake-prone 'subduction' zones off Indonesia and the coast of Pakistan, where one plate of Earth's crust grinds under another. When a big tremor hits, three regional alert centres — in Australia, Indonesia and India — spring into action. Scientists there use seismic data to estimate how much the earthquake has displaced the ocean floor. Then they compare the real quake with model scenarios in which they have calculated what size of tsunami might be produced. The centres alert national governments about what to expect, and data from coastal sea-level gauges and a handful of tsunameters — buoys floating in the open ocean that can detect the passage of large waves — can help to confirm whether a major swell is making its way across the ocean.

What happens next is up to each country — but warnings often fail to travel the 'last mile' to people living in areas, often remote, that are at risk of being swamped. "We really do need to focus on that last mile," says Tony Elliott, head of the warning system's intergovernmental coordination group in Perth, Australia.

In tsunami-prone Indonesia, a German–Indonesian team has worked to develop warning communication chains in 26 provinces and districts. Seven years into the project, only about half of those 26 had implemented a functional warning service that reached all the way down to the local level, says Harald Spahn, a disaster-management consultant formerly with the German development agency GIZ.

And even when alerts do make it to people

at risk, those people do not always behave as disaster experts would wish. In April 2012, a magnitude-8.6 earthquake hit off the coast of Sumatra. Instead of going to shelters, as emergency managers had hoped, many people tried to drive away. The roads in Aceh province became clogged. Fortunately, the geology of that quake meant that it produced only a very minor tsunami.

In Indonesia, Spahn and his colleagues focused on four pilot regions to develop ways to complete the communication chain. They developed tsunami-hazard maps to work out which communities were most at risk. Then they devised a brochure that lays out the warning signs of an approaching tsunami and what to do when one might be on the way. Finally, they helped to develop a three-tier alert system that was adopted at the national level. The tiers depend on the height of the expected tsunami, and specify the action that government officials should take — such as to move people off and away from the beach, evacuate in a limited fashion, or evacuate completely.

Spahn says that the tsunami-alert system can be useful even when no tsunami is coming. In September 2009, a magnitude-7.6 earthquake killed more than 1,100 people in and around the city of Padang on the western coast of Sumatra. The tsunami-alert system indicated that there would be no big wave, which let emergency officials respond more quickly to the earthquake damage.

Maintenance will be key to keeping the information flowing. The IOTWS cost more than $450 million to set up, with most funding coming from Australia, Indonesia and India. If a piece of equipment breaks, it is up to the country that installed it to fix it. The deep-sea buoys in particular are expensive and prone to vandalism or accidental damage from passing ships.

The Indian Ocean countries have varying levels of motivation to keep the system going, Elliott notes. Nations that are farther from likely sources of great earthquakes are less engaged.

Experts say that the best chance of keeping the system operating for the next decade and beyond is to make sure that tsunami alerts are woven into the national fabric for dealing with other kinds of emergency, from cyclones to landslides, many of which use the same sensing networks and communication channels.

"We've done a lot," says Bailey. "We've just got to hang onto it now." ■

→

MORE ONLINE

ERIKA CHECK HAYDEN

SPACE

Europe plans Moon landing

Space–agency scientists propose piggybacking on two Russian missions.

BY ELIZABETH GIBNEY

Science ministers in Europe have resurrected plans to explore the Moon's surface — and the only strategy currently on the table is to join two uncrewed Russian missions. The developments, which follow the shelving of a proposed European Space Agency (ESA) Moon lander two years ago, come amid growing political tensions between Russia and Western nations.

On 2 December, at a meeting in Luxembourg to determine ESA's policy, the space agency got the go-ahead and funding to investigate "participation in robotic missions for the exploration of the Moon". Science ministers from the ESA member states did not approve collaboration with Russia specifically, but at the meeting, ESA scientists presented a proposal to join Russia on its missions to put a lander and a rover on the Moon's south pole.

Money for lunar exploration will come from a pot of €800 million (US$980 million) contributed by ESA's member states and dedicated to international space exploration; the pot will primarily pay for activities on the International Space Station and the development of a propulsion module for NASA's Orion spacecraft, which is eventually designed to carry astronauts to deep space, and was tested on 5 December in an uncrewed space flight (see page 148).

In the 45 years since astronauts first walked on the Moon, no European country or space agency has launched a mission to the Moon's surface. And no lander or astronaut has been to the lunar south pole, a region thought to contain ice and thus deemed a probable spot for any future permanent lunar base. A 12-kilometre-deep crater there might provide access to material from the Moon's interior, also making it attractive for scientific study, says Ian Crawford, a lunar scientist at Birkbeck, University of London. The ancient material could reveal details of the collision between a Mars-sized planet and early Earth that is thought to have produced the Moon. "The idea

"It would be crazy that an agency like ESA would not be part of lunar exploration."
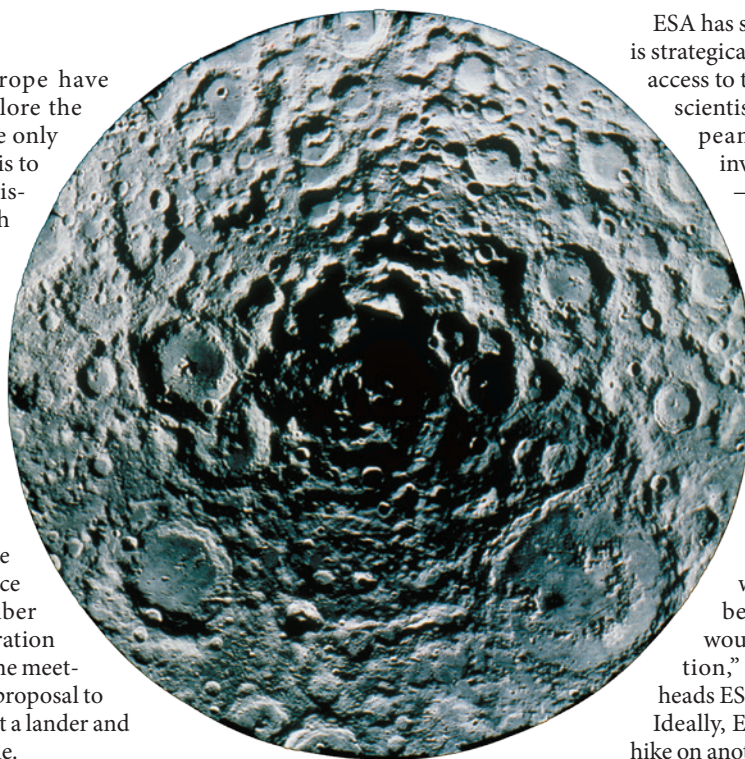
The Moon's south pole is unexplored territory.

that we've 'been there and done that' did last for a long time, but that's gone away now," says Crawford. "The Moon still has a lot to tell us."

A Moon lander proposed by ESA failed to gather enough support at a similar meeting of ministers in 2012. That left European scientists and industry mobilized to go — but without a mission. A group of ESA scientists has been discussing a partnership with the Russian space agency, Roscosmos, ever since.

The group's proposal, aired for the first time at the Luxembourg meeting, is that ESA contribute to Roscosmos's Luna-Resource Lander, also known as Luna 27, which is scheduled for launch in 2019, as well as the Lunar Sample Return, planned for the early 2020s. The first will study the lunar soil and atmosphere at the south pole; the second would bring samples back to Earth. ESA would provide precision landing and communications equipment, as well as drill and analysis instruments.

The ministerial decision, in principle, means that ESA can start to fund efforts to incorporate these technologies into the mission — although whether it will do so has still to be agreed. The preliminary phase is estimated to cost up to €50 million. The total price would be much higher, perhaps in the hundreds of millions.

ESA has said that pursuing lunar missions is strategically important, not only to secure access to the Moon's surface for European scientists, but also to ensure that European expertise and technology is involved in future lunar exploration — including, ultimately, international crewed missions and even a permanent lunar base. NASA currently has no plans to land on the Moon (Orion will be designed to take astronauts into lunar orbit), but Russia, China, Japan and several private companies are making plans to put rovers on the body. Representatives from these nations have more than hinted that permanent Moon bases and human exploration would be the next steps. "It would be crazy that an agency like ESA would not be part of lunar exploration," says Bérengère Houdou, who heads ESA's Lunar Exploration Office.

Ideally, Europe would not need to hitch-hike on another agency's mission to get to the Moon, but the potential Russian collaboration is "a very welcome plan B", says Crawford. "We're primed for a lunar mission, so it's absolutely timely."

It is not clear whether the sour relationship between Russian and Western leadership will affect the proposal's chances of success. Crawford calls it "a potential worry" but stresses that so far, geopolitical problems have not affected space cooperation. ESA officials say that cooperation is continuing normally on existing missions that involve European–Russian collaboration, such as the International Space Station and ExoMars, which will put a demonstration lander on the red planet in 2016 to test technologies for a rover that will land in 2018 to search for signs of past life. The ExoMars rover received the funding it needs to stay on track for 2018 at the 2 December meeting.

In the longer term, Crawford believes that Europe should be looking beyond collaboration with Roscosmos. He adds that China's space agency, which last year became the first since the 1970s to put a lander on the Moon, is the only one that has working scientists and engineers who have Moon-landing experience. "There must be a case," he says, "for ESA broadening its collaboration with other potential space-faring nations." ∎

People with symptoms of Ebola often have to wait days for a diagnosis.

PUBLIC HEALTH

Ebola experts seek to expand testing

Rapid local diagnosis is essential for curbing spread.

BY DECLAN BUTLER

The Ebola crisis in West Africa is approaching the one-year mark, with no clear end in sight. At present, fewer than one in five people with Ebola is diagnosed within two days of becoming infectious, according to the World Health Organization (WHO). Yet in the absence of a safe and effective vaccine, the only way to end the epidemic is to quickly identify and quarantine people who have been infected.

A major problem is that relatively few laboratories in West Africa have the necessary equipment and personnel to test blood samples from people thought to have Ebola (see 'Delayed diagnoses'). But that could soon change. Experts are gathering in Geneva, Switzerland, on 12 December to work out which diagnostic tools could be used wherever Ebola strikes.

The meeting, convened by the WHO and the non-profit Foundation for Innovative New Diagnostics (FIND), also in Geneva, seeks to identify tests that can be used by untrained staff, do not require electricity or can run on batteries or solar power and use reagents that can withstand temperatures of 40 °C. Experts will also discuss how such diagnostics could be rolled out widely in Ebola-stricken areas, and will develop a six-month plan to improve access to testing. If the push succeeds, it would mark an important strategic shift in efforts to end the epidemic.

In addition to reducing Ebola's spread, localized testing of cases would minimize care delays, says Daniel Kelly, an infectious-disease researcher at the University of California, San

Francisco, who has been working in Sierra Leone (see *Nature* **513,** 145; 2014). People die from Ebola when the rapid fluid loss from bleeding, vomiting and diarrhoea causes the heart to stop pumping blood and other organs to fail. "Every second counts," Kelly says. "A faster time to an Ebola diagnosis will save lives."

Most available tests for the virus rely on a technology called reverse-transcriptase polymerase chain reaction (RT-PCR), which detects genetic sequences specific to Ebola in blood, serum and other bodily fluids. These methods are highly sensitive, but require skilled scientists working in sophisticated labs that have high-level biocontainment measures. Access to consistent power supplies and refrigeration is essential, and the tests are expensive, at roughly US$100 apiece.

These requirements put such diagnostic tools out of reach of many hard-hit parts of West Africa, prompting the WHO to establish on 18 September an emergency mechanism for reviewing other, experimental tests. Those that seem promising will be sent to independent laboratories to assess whether they live up to their manufacturers' claims; tests that succeed will be cleared for purchase by the WHO and other United Nations agencies, under a one-year emergency authorization.

PROMISING START

So far, the WHO has received 17 applications from diagnostic companies. Although it has not released the list of candidates, many of the likely contenders and competing technologies are known.

They include 13 RT-PCR tests, many of which have been modified to make them easier to use. Several are at least partially automated: the ones that are easiest to use involve loading a blood sample into the machine, pushing a button and waiting for results that can arrive in as little as an hour. Some of the systems have also been adapted for use in harsh field environments. Mark Perkins, FIND's chief scientific officer, expects that the WHO will approve some of these tests early next year.

The other four candidates are tests that detect antigens to Ebola in blood and other fluids — in many cases, using the same strip format and analytical technique, called enzyme-linked immunosorbent assay (ELISA), as over-the-counter pregnancy-test kits. Such tests are cheap to mass produce, do not require electricity or refrigeration and use just a drop of blood.

There are some potential drawbacks, however. The strip-format ELISA used in pregnancy-test-like diagnostics tend to be several orders of magnitude less sensitive than RT-PCR, so may not be able to detect Ebola just after symptoms appear, says Sterghios Moschos, an industrial biotechnology researcher at the University of Westminster, UK. (Moschos is developing a rapid RT-PCR test called EbolaCheck.) Other antigen-based
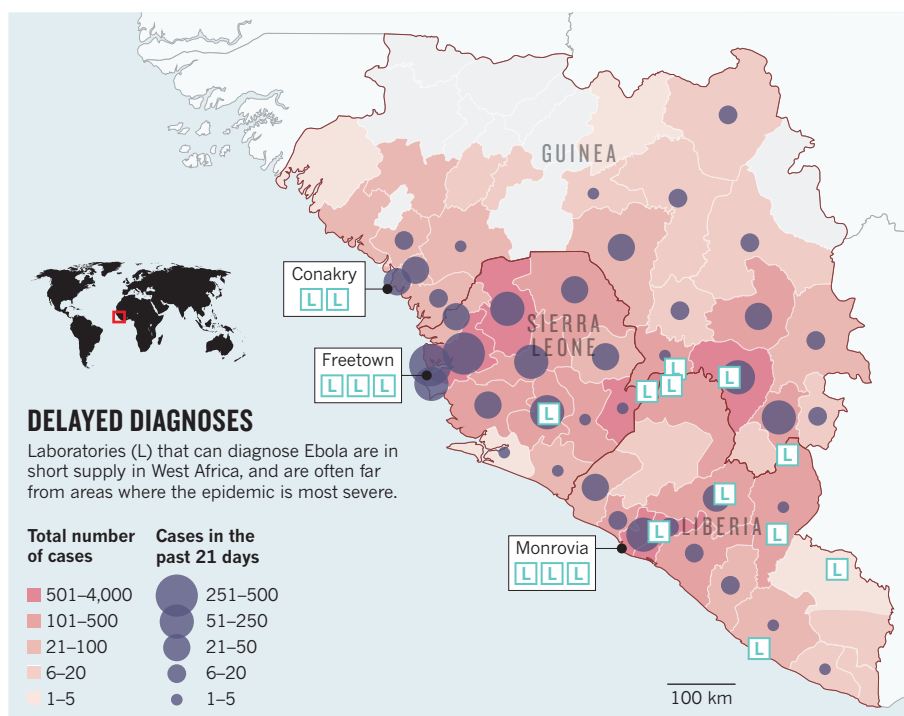
tests use different technologies that are likely to detect Ebola early in its course.

The accuracy of antigen-based Ebola tests is of particular concern in Africa, because individuals there often carry antigens to several viruses and parasites, such as those that cause malaria, tuberculosis and hepatitis, which might muddy the results for Ebola. Tests that work well in the lab against blood and serum that have been artificially spiked with Ebola may not cope as well with clinical samples gathered in the field.

But Robert Garry, a virologist at Tulane University in New Orleans, Louisiana, who is developing an antigen-based Ebola strip test, is nonetheless confident about that approach after successfully developing a similar diagnostic for Lassa fever. Garry, who is working with diagnostics firm Corgenix of Bloomfield, Colorado, has just completed initial field trials of his Ebola test in Sierra Leone. "It's looking very good," he says.

Perkins is cautious. Until all the new Ebola tests are tried in the field, and independently evaluated by the WHO, "the jury is still out", he says. ∎

DELAYED DIAGNOSES

Laboratories (L) that can diagnose Ebola are in short supply in West Africa, and are often far from areas where the epidemic is most severe.

Total number of cases	Cases in the past 21 days
501–4,000	251–500
101–500	51–250
21–100	21–50
6–20	6–20
1–5	1–5

100 km

A report from the front line

Nature reporter Erika Check Hayden is in Sierra Leone tracking the Ebola epidemic. More of her dispatches can be found at **www.bit.ly/eboladiary**.

1 DECEMBER: MIXED SIGNALS
Arriving at an Ebola treatment centre outside Sierra Leone's capital, Freetown, I heard celebratory singing and clapping: three survivors of the disease were preparing to leave. Staff at the centre in Kerry Town, which is run by the non-profit organization Save the Children, presented the survivors with laminated certificates documenting their Ebola-free status.

Almost a year after the first Ebola cases were reported, there are signs of hope — such as these survivors. But the number of cases is still rising in some areas in Sierra Leone, including Freetown, and there are still not enough treatment beds for everyone. There is no single reason why the epidemic is still growing in parts of Sierra Leone, but a contributing factor is the difficulty of convincing people who have never previously experienced the disease to change the way that they live, care for the sick and bury the dead.

3 DECEMBER: CLOSE TO HOME
"From the back, from the back!" shouts Halima Shyllon, the nurse matron of a newly opened Ebola treatment centre in Makeni.

She and two workers are supervising doctor Moges Tadesse as he removes the hooded Tyvek suit that he has been wearing to treat patients for the past hour.

Before this treatment centre opened on 24 November, people with Ebola were sent elsewhere in the country, to wherever there was a bed available. Their families often never saw or heard from them again, because many died in remote care facilities.

The Makeni centre, the district's first, is run by the African Union and is largely staffed by Africans: Shyllon works for Sierra Leone's health ministry, and Tadesse is Ethiopian. There are Ugandan doctors on site and Nigerian workers will arrive soon.

One patient, Usman Fofanah, was so sick when he arrived in Makeni a few days ago that he does not remember the journey from Port Loko, about 80 kilometres away. Fofanah has lost his grandmother, an aunt and two sisters to Ebola. Still, he is smiling: he feels much better and yesterday he was able to speak by phone to his mother, who had feared him dead.

7 DECEMBER: IN QUARANTINE
When we arrive at the school in Tambiama, in the Bombali district of northern Sierra Leone, a few people stand in a dirt yard behind a strip of red and white quarantine tape. As a policeman and a soldier call to the rest, the men, women and children slowly file out

— the friends and neighbours of a woman who died from Ebola on 14 November.

When a local priest asks how they feel, the 40 or so villagers all say that they are fine; a few even break into a spontaneous dance to prove it. If all goes well, they will be free in a matter of days. But they are restless. "They are not feeling good at all — there [is] no free movement," says Shekub Mansary, a health worker who is translating for James Koroma, a former teacher at the primary school.

Quarantines have been widely deployed in this outbreak, but they are a crude tool. Bombali district has been under quarantine since September; only vehicles with special permissions can enter or leave. But on a recent day (6 December), there were 11 new Ebola cases. And although 1 million Sierra Leoneans now live in quarantined districts, Ebola is still infecting new areas.

People resent, and sometimes resist, the restrictions. A few weeks ago, villagers near Tambiama fought off quarantine officers with machetes.

Early on, there were problems guaranteeing that quarantined families in many areas had enough to eat. Such conditions make opposition to quarantine understandable, says Catherine Bolten, an anthropologist at the University of Notre Dame in Indiana. "You might or might not get Ebola, but you'd definitely know if you're starving to death." ∎

MEDICINE

T-cell therapy extends cancer survival to years

Firms embrace immunotherapy to fight intractable leukaemias and lymphomas.

BY HEIDI LEDFORD

When immunologist Michel Sadelain launched his first trial of genetically engineered, cancer-fighting T cells in 2007, he struggled to find patients willing to participate. Studies in mice suggested that the approach — isolating and engineering some of a patient's T cells to recognize cancer and then injecting them back — could work. But Sadelain did not blame colleagues for refusing to refer patients. "It does sound like science fiction," he says. "I've been thinking about this for 25 years, and I still say to myself, 'What a crazy idea.'"

Since then, early results from Sadelain's and other groups have shown that his 'crazy idea' can wipe out all signs of leukaemia in some patients for whom conventional treatment has failed. And today, his group at the Memorial Sloan Kettering Cancer Center in New York City struggles to accommodate the many people who ask to be included in trials of the therapy, known as adoptive T-cell transfer.

At the American Society of Hematology (ASH) meeting held in San Francisco, California, on 6–9 December, attendees heard dozens of talks and poster presentations on the promise of engineered T cells — commonly called CAR (chimaeric antigen receptor) T cells — for treating leukaemias and lymphomas. The field has been marred by concerns over safety, the difficulties of manufacturing personalized T-cell therapies on a large scale, and how regulators will view the unusual and complicated treatment. But

those fears have been quelled for some former sceptics by data showing years of survival in patients who once had just months to live.

"The numbers are pretty stunning," says Joseph Hedden, an analyst for the London-based market-research firm Datamonitor Healthcare. "Companies have clearly decided that it's worth the pitfalls of how much this therapy is going to cost to develop."

At least five major pharmaceutical companies have invested in developing CAR-T-cell therapy over the past three years. Such interest from industry is a dramatic turn for a field that once consisted of a handful of academic medical centres. Small biotechnology firms have also sprung up to develop CAR T cells, including Kite Pharmaceuticals of Santa Monica, California, which raised US$127.5 million when it went public in June. And investors pumped $310 million into another CAR-T-cell company, Juno Therapeutics of Seattle, Washington, this year. "There is no doubt there has been a shift," says Juno chief executive Hans Bishop.

Most of these efforts focus on killing the cancerous, antibody-producing B cells behind some leukaemias and lymphomas. Researchers do this by engineering T cells to recognize a protein on the surface of most B cells — CD19 — and attacking cells that display it (see 'Call to arms'). Finding proteins that are expressed only on cancer cells can be difficult, and CD19 represents a compromise: the treatment sometimes wipes out all B cells, cancerous and healthy alike, but patients can survive without them.

At the ASH meeting, Sadelain and his colleagues reported that this approach left no signs of cancer in all six patients with lymphoma who were enrolled in one trial. In another presentation, immunologist Carl June of the University of Pennsylvania in Philadelphia showed that targeting CD19 reduced cancer burden in 9 of 23 patients with chronic lymphocytic leukaemia. In a more aggressive disease called acute lymphoblastic leukaemia, 27 of 30 patients had no signs of cancer after therapy and the CAR T cells remained in their blood two years later.

But studies also highlight the risks of revving up immune responses. In April, at least five CAR-T-cell trials were halted after a series of patient deaths linked to unusually high levels of a protein called interleukin-6, which promotes inflammation, as well as other inflammatory molecules. Interleukin-6 is part of the body's normal response to infection. But the intense immune onslaught launched by CAR T cells can send interleukin-6 levels soaring. The trials resumed after investigators adjusted their protocols to better monitor and treat the problem.
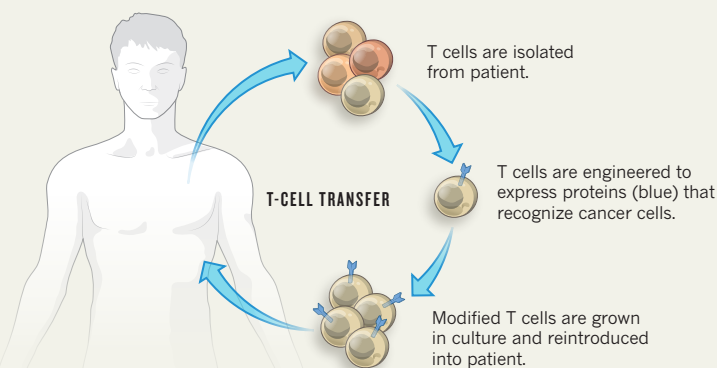
These safety risks, as well as the difficulty of manufacturing CAR T cells, are still putting many drug companies off, says Andrew Baum, the London-based head of global health-care research for Citi, an investment bank headquartered in New York City. "The bulk of the multinationals are standing back and watching, rather than getting engaged here," he says.

When CAR T cells do reach the market, they will not be cheap. Baum says that some sponsors are tentatively planning to price their therapies higher than bone-marrow transplants, which can exceed $500,000. The cost may be so high, he says, that companies are forced to set up a reimbursement scheme in which they are paid only when a patient benefits from the treatment. Baum estimates that peak sales of CAR-T-cell therapies will reach $10 billion annually, although that amount will depend on what competing therapies emerge and whether the treatment can be extended to other cancers.

For now, Sadelain, a scientific founder of Juno Therapeutics, hopes that the attention from industry will spur the field. He remembers his postdoc days, when he struggled to insert genes into T cells and colleagues asked him why he was bothering. "We've never had this kind of investment in the field before," he says. "It's hard to believe — sometimes I still pinch myself." ∎

CALL TO ARMS

A promising cancer therapy called adoptive T-cell transfer genetically engineers a patient's own immune cells to target tumours.

T-CELL TRANSFER

T cells are isolated from patient.

T cells are engineered to express proteins (blue) that recognize cancer cells.

Modified T cells are grown in culture and reintroduced into patient.

Paul Allen's latest philanthropic endeavour will be modelled on his successful brain institute.

SYSTEMS BIOLOGY

Microsoft billionaire takes on cell biology

New Allen institute will study and simulate cell behaviour.

BY EWEN CALLAWAY

Billionaire businessman and philanthropist Paul Allen plans to pump US$100 million into investigating the most basic unit of life — the cell.

The Allen Institute for Cell Science, which was launched on 8 December, will be modelled on the Microsoft co-founder's Allen Institute for Brain Science in Seattle, Washington, which since 2003 has spent hundreds of millions of dollars creating a series of 'brain atlases' that have become go-to portals for neuroscientists interested in where particular genes are active or how distant neurons communicate.

As its first project, the new Allen institute will develop an analogous 'cell observatory' that will display how a cell's working parts, such as ribosomes, microtubules and mitochondria, interact and operate over time, says executive director Rick Horwitz. He has shuttered his cell-biology laboratory at the University of Virginia in Charlottesville to lead the institute in Seattle, Washington. The 70 or so scientific staff who will join the institute will work together on the overall goals of the observatory — to build a global view of the myriad activities inside cells — rather than on their own interests. "It's going to be much more like the Manhattan Project," Horwitz says.

Mapping every little detail of every kind of cell is a tall order, even with the backing of the

world's 27th richest person. "Our problem is that this thing could blow up on us. It could be very, very big," Horwitz says. "We're going to make judicious decisions to try to contain it."

Some of those choices have already been made, after meetings this year with leading cell biologists. The institute will study human induced pluripotent stem cells (cells coaxed into an embryonic stem-cell-like state) as they differentiate in the lab into two cell types: heart-muscle cells called cardiomyocytes; and the epithelial cells that line body cavities. These tissues were chosen as much for their relevance to

disease — cardiomyocytes malfunction in heart disease and most cancers arise in epithelial tissues — as for the ease with which they can be reproducibly generated and grown in the lab.

The institute's plan is to engineer many different cell lines and determine how different cellular components respond to stimuli such as infection or exposure to a drug. These data will then guide the construction of computer models aimed at predicting how cells operate under various conditions, and all the information gained will be made available online. The institute will also distribute its cell lines so that other scientists can build on its work.

The $100 million is set to cover the first five years, after which Allen will review the observatory's achievements and decide whether to keep on funding it, Horwitz says. The Allen brain institute was also started with $100 million and has received subsequent funding of $400 million. Allan Jones, chief executive of the brain institute, says that the cell institute's success will be measured both in terms of research output — the brain atlases have yielded dozens of papers — and its broader impact on biology. "You need to make a high-quality product that people trust and believe in," he says.

Just as many neuroscience studies now begin with a trawl through the Allen institute's brain atlases, the cell observatory "will be the place cell biologists go to start projects", says Sandra Schmid, a member of the cell institute's advisory board and a cell biologist at the University of Texas Southwestern Medical Center in Dallas. Ruedi Aebersold, a systems biologist at the Swiss Federal Institute of Technology in Zurich, is enthusiastic about the plans, but says that it will take time to see whether the institute leaves an indelible mark on cell biology. "One would want to ask eventually, in five years, how this effort has accelerated that research," he says.

Trey Ideker, a systems biologist at the University of California, San Diego, says predicting how cells behave is an exciting, if ambitious, goal. "My concern is that they need focus," he says. "I think Rick's mandate is he's got to tell the world what the goal of this institute is." ∎

CORRECTIONS

The *y*-axis on the graphic in the News story 'US–China climate deal raises hopes for Lima talks' (*Nature* **515,** 473–474; 2014) was out by a factor of 10. It should have been 0–35 gigatonnes not 0–3.5 Gt.

The picture caption in the story 'Ocean observatory project hits rough water' (*Nature* **515,** 474–475; 2014) gave the wrong date for the completion of the Ocean Observatories Initiative network: it will finish in May 2015, not March.

The story 'Green List promotes conservation hotspots' (*Nature* **515,** 322; 2014) misstated why original inhabitants of the Chagos Islands cannot return: it is owing to policies of the

British Indian Ocean Territory administration.

The article 'Rival species recast significance of 'first bird'' (*Nature* **516,** 18–19; 2014) incorrectly referred to '*Microraptor xui*' instead of '*Microraptor gui*'. It also failed to attribute the *Archaeopteryx* silhouette in the graphic to Vladimir Nikolov.

The story 'Climate tinkerers thrash out a plan' (*Nature* **516,** 20–21; 2014) incorrectly stated that discussions at the meeting would feed into a report that the US National Academies intends to release early next year. And the caption stated that the futuristic device would spray sea water into the stratosphere. Actually, the lower atmosphere is the target.

Life
– *a status report*

SPECIES ARE DISAPPEARING QUICKLY — BUT RESEARCHERS ARE STRUGGLING TO ASSESS HOW BAD THE PROBLEM IS.

BY RICHARD MONASTERSKY

Of all the species that have populated Earth at some time over the past 3.5-billion years, more than 95% have vanished — many of them in spectacular die-offs called mass extinctions. On that much, researchers can generally agree. Yet when it comes to taking stock of how much life exists today — and how quickly it will vanish in the future — uncertainty prevails.

Studies that try to tally the number of species of animals, plants and fungi alive right now produce estimates that swing from less than 2 million to more than 50 million. The problem is that researchers have so far sampled only a sliver of Earth's biodiversity, and most of the unknown groups inhabit small regions of the world, often in habitats that are rapidly being destroyed.

The International Union for Conservation of Nature (IUCN) highlighted the uncertainty in the latest version of its Red List of Threatened Species, which was released in November. The report evaluated more than 76,000 species, a big increase over earlier editions. But that is just 4% of the more than 1.7 million species that have been described by scientists, making it impossible to offer any reliable threat level for groups that have not been adequately assessed, such as fish, reptiles and insects.

Recognizing these caveats, *Nature* pulled together the most reliable available data to provide a graphic status report of life on Earth (see 'Life under threat'). Among the groups that can be assessed, amphibians stand out as the most imperilled: 41% face the threat of extinction, in part because of devastating epidemics caused by chytrid fungi. Large fractions of mammals and birds face significant threats because of habitat loss and degradation, as well as activities such as hunting.

Looking forward, the picture gets less certain. The effects of climate change, which are hard to forecast in terms of pace and pattern, will probably accelerate extinctions in as-yet unknown ways. One simple way to project into the future would be to assume that the rate of extinction will be constant; it is currently estimated to range from 0.01% to 0.7% of all existing species a year. "There is a huge uncertainty in projecting future extinction rates," says Henrique Pereira, an ecologist at the German Centre for Integrative Biodiversity Research in Leipzig.

At the upper rate, thousands of species are disappearing each year. If that trend continues, it could lead to a mass extinction — defined as a loss of 75% of species — over the next few centuries.

Conservation policies could slow extinctions, but current trends do not give much comfort. Although nations are expanding the number of land and ocean areas that they set aside for protection, most measures of biodiversity show that pressures on species are increasing. "In general, the state of biodiversity is worsening, in many cases significantly," says Derek Tittensor, a marine ecologist with the United Nations Environment Programme's World Conservation Monitoring Centre in Cambridge, UK.

Despite all the uncertainty, researchers agree that they need to devote more attention to evaluating current and future risks to biodiversity. One approach is to develop comprehensive computer models that can forecast how human activities will alter ecosystems. These general ecosystem models, or GEMs, are in their infancy: earlier this year, Tittensor and his colleagues published initial results from the first global model that seeks to mimic all the major ecological interactions on Earth in much the same way as climate models simulate the atmosphere and oceans (M. B. J. Harfoot *et al. PLoS Biol.* **12,** e1001841; 2014). Building the GEM took 3 years, in part because the model tries to represent all organisms with body masses ranging from 10 micrograms (about the weight of small plankton) to 150,000 kilograms (roughly the size of a blue whale). "It needs a lot more development and testing, and ideally there will be a lot more variety of these models," says Tittensor. But if they do a decent job of capturing the breadth of life in a computer, he says, "they have real potential to alert us to potential problems we wouldn't otherwise detect". ■ SEE EDITORIAL P.144

NATURE.COM
For an interactive version of the graphic, visit:
go.nature.com/x8w3ec

Richard Monastersky *is an editor with* Nature *in Washington DC.*

Life under threat

Thousands of species are currently deemed to be threatened, but the true number of species at risk of extinction may be much higher. Estimates suggest that between 500 and 36,000 species might be disappearing each year. The best data are for well-studied groups — mammals, birds and amphibians. Much less is known about threats to other groups, such as insects and fish.

ALREADY EXTINCT

TOTAL DOCUMENTED SINCE 1500

79 **145** **36** **505**

- Mammals
- Birds
- Amphibians
- Others

TOTAL
765
EXTINCT
SPECIES
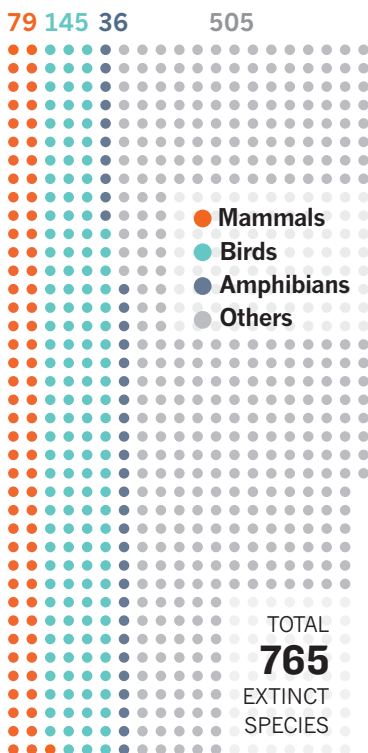
Mammals
1,199
THREATENED SPECIES
26% of described species

Birds
1,373
THREATENED SPECIES
13% of described species

Spheniscus demersus

Burramys parvus

CURRENTLY THREATENED

TOTAL
5,522
THREATENED MAMMALS, BIRDS, AMPHIBIANS AND INSECTS
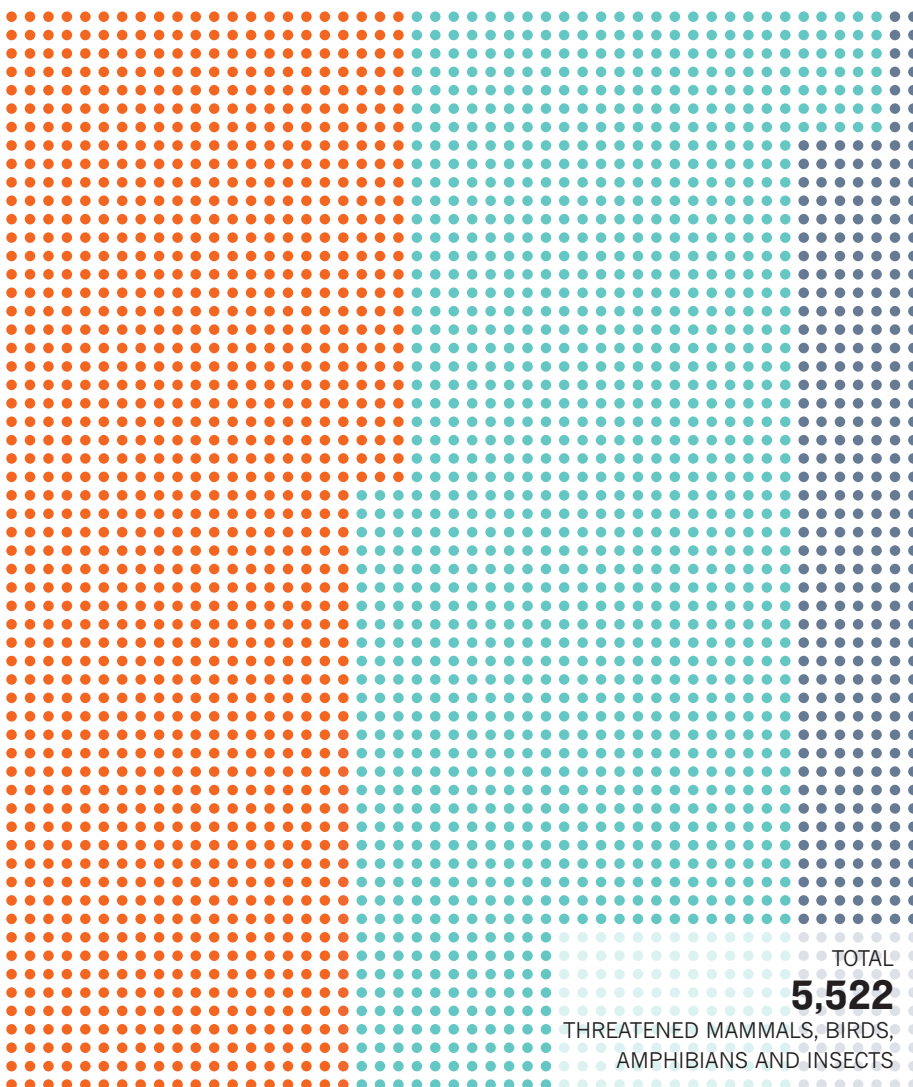
March towards mass extinction

Mass extinctions — loss of 75% of existing species — have happened 5 times in the planet's history. If there are 5 million animal species and they are disappearing at rate of 0.72% per year (the upper end of estimates), a sixth mass extinction could happen by the year 2200. At the low end of the estimated range, a mass extinction would not happen for thousands of years.

EXTINCTIONS PER WEEK

LOWER ESTIMATE
10

HIGHER ESTIMATE
690

BY RICHARD MONASTERSKY | GRAPHIC BY 5W INFOGRAPHIC

Amphibians
1,957
THREATENED SPECIES
41% of described species

Ranitomeya summersi

Insects
993
THREATENED SPECIES
(Only 0.5% of roughly 1 million described have been evaluated. Number of living species may exceed 5 million)

Nicrophorus americanus
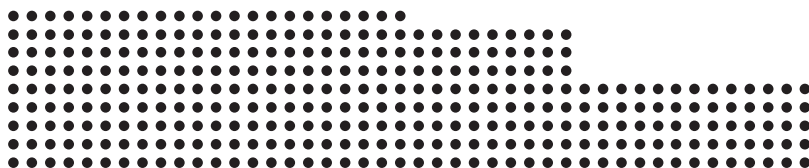
How many species are there?

Estimates of the number of species of animals, fungi and plants vary significantly. That uncertainty clouds understanding of how many species are threatened and how many are going extinct.

ANIMALS

2 million to 11 million predicted species

1,371,500 described species

FUNGI

600,000 to 10 million predicted

48,500 described

307,700 to 450,000 predicted

PLANTS

307,700 described
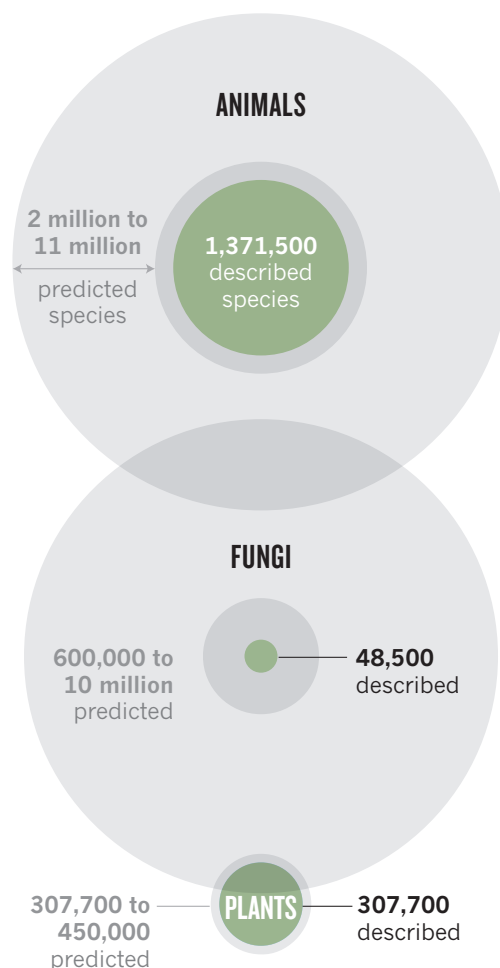
Main threats

Hunting, fishing and other forms of exploitation are a major factor in declines in animal populations, according to the Living Planet Index. Habitat degradation and loss are also dominant threats. Climate change is expected to become a bigger factor over time.

Exploitation **37%**

Habitat degradation and change **31%**

Habitat loss **13%**

Climate change **7%**

Invasive species **5%**

Pollution **4%**

Disease **2%**
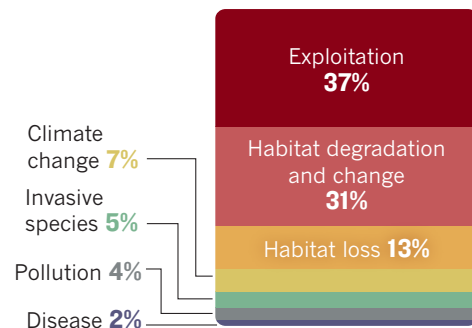
FIGURES HAVE BEEN ROUNDED

SOURCES: Already Extinct, Currently threatened: IUCN Red List. How many species are there?: S. L. Pimm *et al. Science* **344,** 1246752 (2014); B. R. Scheffers *et al. Trends Ecol. Evol.* **27,** 501–510 (2012); IUCN Red List. March towards mass extinction: Pimm *et al.*; C. Mora *et al. Science* **341,** 237 (2013). Main threats: *WWF Living Planet Report 2014.*
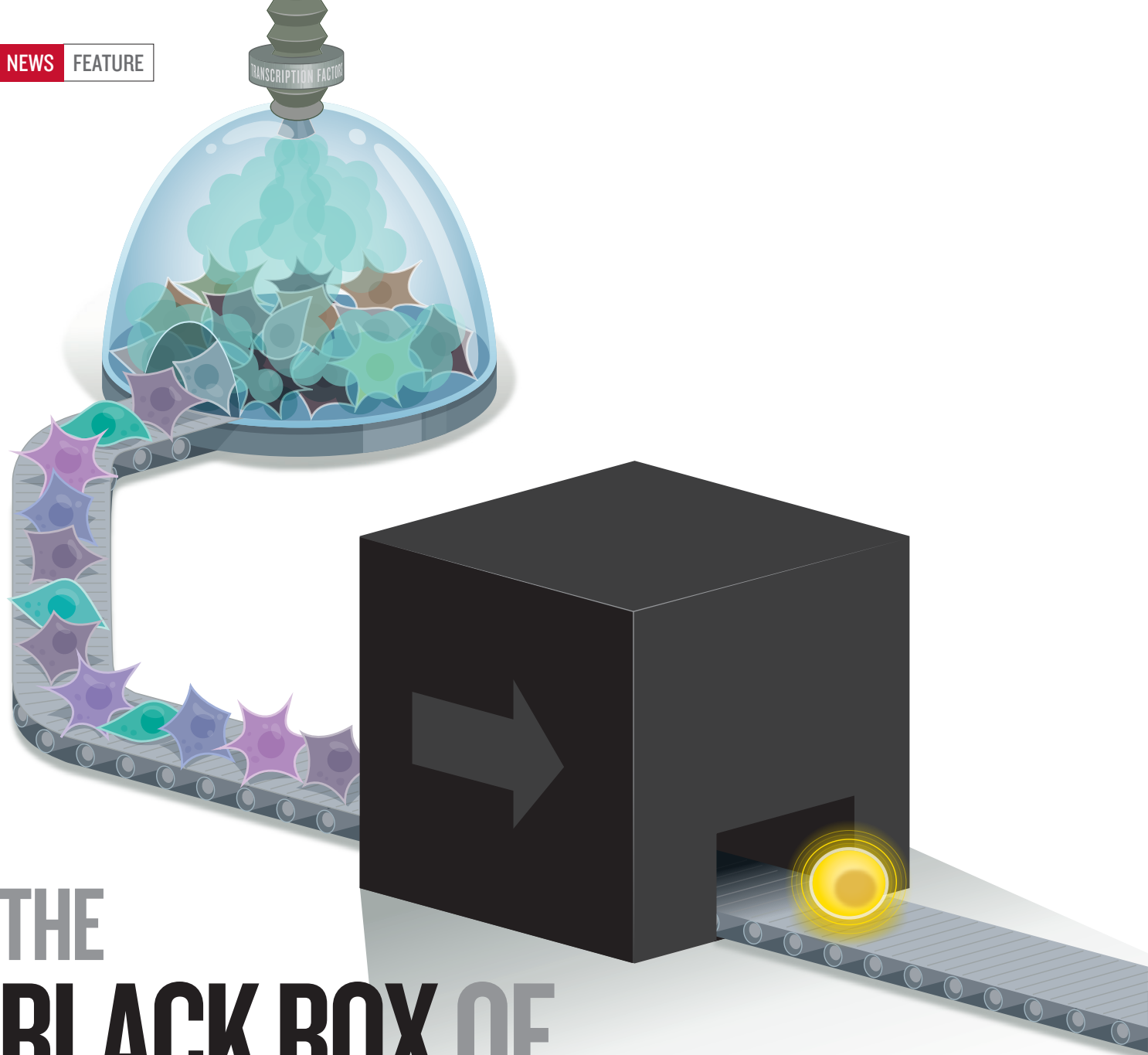
THE
BLACK BOX OF
REPROGRAMMING

BY DAVID CYRANOSKI

Scientists have been reprogramming adult cells into embryonic ones for decades — but they are only now getting to grips with the mechanics.

Eggs and sperm do it when they combine to make an embryo. John Gurdon did it in the 1960s, when he used intestinal cells from tadpoles to generate genetically identical frogs. Ian Wilmut did it too, when he used an adult mammalian cell to make Dolly the sheep in 1996. Reprogramming — reverting differentiated cells back to an embryonic state, with the extraordinary ability to create all the cells in the body — has been going on for a very long time.

Scientific interest in reprogramming rocketed after 2006, when scientists showed that

adult mouse cells could be reprogrammed by the introduction of just four genes, creating what they called induced pluripotent stem (iPS) cells[1]. The method was simple enough for almost any lab to attempt, and now it accounts for more than a thousand papers per year. The hope is that pluripotent cells could be used to repair damaged or diseased tissue — something that moved closer to reality this year, when retinal cells derived from iPS cells were transplanted into a woman with eye disease, marking the first time that reprogrammed cells were

transplanted into humans (see *Nature* http://doi.org/xhz; 2014).

There is just one hitch. No one, not even the dozen or so groups of scientists who intensively study reprogramming, knows how it happens. They understand that differentiated cells go in, and pluripotent cells come out the other end, but what happens in between is one of biology's impenetrable black boxes. "We're throwing everything we've got at it," says molecular biologist Knut Woltjen of the Center for iPS Cell Research and Application at Kyoto University in Japan. "It's still a really confusing process. It's very complicated, what we're doing."

One of the problems, stem-cell biologists say, is that their starting population contains a mix of cells, each in a slightly different molecular state. And the process for making iPS cells is currently inefficient and variable: only a tiny fraction end up fully reprogrammed and even these may differ from one another in subtle but important ways. What is more, the path to reprogramming may vary depending on the conditions under which cells are being grown, and from one lab to the next. This makes it difficult to compare experimental results, and it raises safety concerns should a mix of poorly characterized cells be used in the clinic.

But new techniques are starting to clarify the picture. By carrying out meticulous analyses of single cells and amassing reams of detailed molecular data, biologists are identifying a number of essential events that take place en route to a reprogrammed state. This week, the biggest such project — an international collaboration audaciously called Project Grandiose — unveiled its results[2–6]. The scientists involved used a battery of tests to take fine-scale snapshots of every stage of reprogramming — and in the process, revealed an alternative state of pluripotency. "It was the first high-resolution analysis of change in cell state over time," says Andras Nagy, a stem-cell biologist at Mount Sinai Hospital in Toronto, Canada, who led the project. "I'm not shy about saying grandiose."

But there is more to do if scientists want to control the process well enough to generate therapeutic cells with ease. "Yes, we can make iPS cells and yes we can differentiate them, but I think we feel that we do not control them enough" says Jacob Hanna, a stem-cell biologist at the Weizmann Institute of Science in Rehovot, Israel. "Controlling cell behaviour at will is very cool. And the way to do it is to understand their molecular biology with great detail."

NUCLEAR TRANSFER

When Gurdon and Wilmut reprogrammed frog and sheep cells, respectively, they did it by transferring a differentiated nucleus into an egg stripped of its own DNA. Scientists knew that something in the egg was able to reprogram the nucleus, such that the genes associated with being a skin cell, for example, were switched off and those associated with pluripotency were switched on and triggered a cascade of downstream events. In the following decade, researchers found various new ways to reprogram — adding nuclei to fertilized eggs and to embryonic stem cells — but these methods did little to clarify what it was in the cells that did the reprogramming and how the process worked.

That changed when Shinya Yamanaka and Kazutoshi Takahashi at Kyoto University made iPS cells[1]. They showed that just four proteins that are usually expressed in early embryos or in embryonic stem cells could reprogram an adult cell — and, crucially, they also provided a tool that researchers could use to study reprogramming in a culture dish, something they have been doing ever since. Stem-cell biologists now

that only genes relevant for a fibroblast are expressed. It wouldn't do for a skin cell to suddenly behave like a dividing stem cell, because that can be the route to diseases such as cancer.

Scientists now have a good grip on what happens during the first 48 hours as the four Yamanaka factors, with brute force, kick cells out of this state. In embryonic stem cells, these proteins activate genes in a 'pluripotency network' that keeps cells proliferating indefinitely. But the factors act differently when shoved into a differentiated cell such as a fibroblast. When cell biologist Ken Zaret at the University of Pennsylvania in Philadelphia mapped the location of these factors during the first two days of

"THE ONE THING THAT WE KNOW IS THAT IT'S NOT MAGIC, THERE IS A MECHANISM."

know that after introducing these proteins — sometimes known as the Yamanaka factors — there is a flurry of intense and mostly predictable gene expression. But then, after a few days, the cells enter a mysterious state in which they are dividing but stalled, failing to reprogram further. After a week or so, a slim few — only one in a thousand — become true pluripotent cells[7].

This process is unpredictable, in the sense that it is impossible to know at the beginning which cells will reprogram, and it takes them a long time. But it is predictable in some ways. "Researchers doing it in Germany, Japan and the US will all get the iPS cells about the same time and at about the same rate," says Alexander Meissner at Harvard University in Cambridge, Massachusetts. "The one thing we know is that it's not magic, there is a mechanism. That's good news — we should be able to find it." And yet, Meissner says, it is "almost disappointing" how little progress there is from year to year.

From the cell's point of view, it is an immense task to overcome a fully differentiated state, which is like being in biological lock-down. Take fibroblasts, for example, the connective-tissue cells that scientists often extract from skin and try to reprogram. In the long process by which they gained their identity, these cells' DNA has been stamped with 'epigenetic' markers, chemical modifications such as the addition of methyl groups or changes to the histone proteins that package up DNA. These ensure

reprogramming in human fibroblasts, he found that they were "physically blocked" from reaching their usual target genes by the conformation of the chromosomes[8].

Instead, the proteins head for accessible areas of the chromosomes. Sometimes, they activate genes that force the cell to commit suicide; in others, they bind to distant control regions called enhancers that encourage the activation of genes known to be involved in the reprogramming process. Rudolf Jaenisch, a stem-cell scientist at the Massachusetts Institute of Technology in Cambridge, has labelled this widespread binding of the Yamanaka factors as "promiscuous"[9].

Other studies have illuminated the sweeping changes that take place on chromosomes during this early phase. In a study published in 2011, Meissner's group showed that a type of histone modification that boosts gene expression, called H3K4me2, changes at more than 1,000 positions in the genome of these cells: it was added at many sites on pluripotency genes, and dropped from sites where genes specific for fibroblasts reside[10]. At the same time, the cells look and behave differently: they compact and move around less.

"Our early thought was that the factors create complete chaos," says Meissner. "But this first step is predictable and consistent across all cell types." Now he can almost foretell for a given cell type "which sites might become open to active transcription, which might be modified, and which will stay silent", he says. "That part you can predict. But that doesn't answer the question of what happens next."

↪ NATURE.COM
Listen to a podcast about reprogramming at
go.nature.com/xydbge

The week-long lag that follows flummoxes scientists. The cells soldier on, and some express new genes, but not in a predictable or comprehensible way. Even the H3K4me2 modifications mapped by Meissner do not seem to boost gene expression until much later in the process. "Most cells reach a partially reprogrammed state. Some get beyond that, and we're not sure why," says Meissner. "That is the black box." If a cell starts to pump out Sox-2 protein, however, that is a really good sign that it is progressing. "Once Sox-2 comes on, everything falls in line," says Jaenisch, who studied the activity of nearly 50 genes in individual cells as they went through reprogramming[11]. Within a few days, the production of this and other transcription factors necessary for pluripotency all ramp up.

But why does all this take so long, and why is it so rare? "We don't understand why it can't be faster," says Woltjen. He suggests that a cell might need to go through several divisions, each taking at least half a day, to reshape its epigenetic state. "Perhaps that's one limiting factor," he says.

Yamanaka offers several possible explanations for the low conversion rate. One is that

idea that variability in the reprogramming process is producing fundamentally different cells. The project, launched in 2010 by some 30 senior scientists at 8 research institutes, was motivated by Nagy's desire to open up the black box. "I wanted to find out what was in it," he says. After triggering reprogramming with the Yamanaka factors, the team collected 100 million cells per day for a month, and then regularly analysed their production of protein and RNA, their changing methylation state and more. The methylation analyses alone produced so much data that collaborators resorted to sharing it on terabyte hard drives that they FedEx-ed around the world. The size of the undertaking also inspired the project's title, Nagy says. "The name just came out of my head when I was considering how much data was being collected," he says.

A CLASS OF ITS OWN
The headline finding is the new category of pluripotent cell, called F-class cells after the fuzzy appearance of the cell colonies. These cells were produced with a small tweak to the iPS-cell recipe: instead of stopping expression

true. According to this 'stochastic' model, as the reprogramming factors trigger cascades of molecules, some cells will drift into a reprogrammed state and some will not, and which way they go cannot be predicted.

But some studies, including one by Hanna[12], show that the reprogramming method can be tweaked to make the process more efficient — suggesting that the randomness can be controlled or even eliminated. These studies imply that reprogramming can be switched from a stochastic process to a deterministic one, in which one step inevitably follows the next to a new cell state.

Many scientists now say that reprogramming involves both deterministic phases — at the start and end — and a stochastic phase, which is the mysterious week in the middle. Hanna plays down the debate altogether, seeing little contradiction between the two sides. "I do not believe there is a stochastic versus deterministic camp." He compares reprogramming to flipping a coin: each flip will have a random outcome, but after 100 flips, close to 50% of them will have come up heads. Similarly, whether a given cell flips into a reprogrammed state might be random. But over time, a reprogramming method should produce a certain percentage — maybe 10% — of pluripotent cells every time. Further experiments might resolve the debate, says Zaret, by pinpointing the events that snap the cells out of their week-long lethargy.

For Zaret, the reprogramming debate offers a window on a bigger concept: how order in biology arises from randomness. "Cellular systems are built upon intrinsic noise and stochastic events that somehow elicit cell fates that are locked down and do not switch back and forth," he says. This question is at the basis of cell type control, he says, and draws him to the research.

For others, like Yamanaka, the incentive to open the black box is a practical one. More-efficient reprogramming makes for better experiments and a more reliable source of cells that can eventually be used in human medicine. "The motivation of my research is to treat patients," he says. "Anything that helps push iPS cells into the clinic excites me." ■

"I'M NOT SHY ABOUT SAYING GRANDIOSE."

the starting cell population is a rainbow of cell types. The chunk of tissue used to derive fibroblasts, for example, probably contained a mix of subtly different cell types; even those that are fibroblasts will differ slightly in the blend of proteins and other molecules they contain. Furthermore, cells growing in culture are constantly shuttling back and forth between different states. This means that the introduced reprogramming factors will affect each cell differently. "What works for one subset of the population will not work for others," Yamanaka says. Minor differences in cell culture and the relationship with neighbouring cells also make it difficult to control all the variables and command the cells like an obedient army, he adds. "A perfect implementation is impossible."

Researchers are now trying to classify some of the cell types that come out of the black box, and are tinkering with reprogramming techniques to see if they can pin down how and where they diverge. Woltjen, for example, has shown that the ratio of the different reprogramming factors affects the type of cells produced. One set of conditions has a high success rate, but the resulting cells end up in a partially reprogrammed, unstable state; another has a low efficiency but produces mainly high-quality iPS cells.

Project Grandiose has also supported the

of the reprogramming factors after a few days, the researchers continued to supply them. "That leads to a bifurcation," says Nagy.

F-class cells are different from iPS cells because they fail one of the most stringent tests of pluripotency: when injected into mouse embryos they cannot contribute to tissues in the resulting chimaeric mice. For this reason, some critics say that F-class cells could be what other scientists have been calling 'partially reprogrammed' cells. But Nagy says that cells do not have to contribute to chimaeras to be considered pluripotent, and points to the cells' other characteristics of pluripotency: for example, they form what is known as a teratoma, which contains a range of differentiated cell types.

Nagy says that others have overlooked the F-class state because they were only looking for cells that were similar to embryonic stem cells, whereas his team was "unbiased by expectation of what pluripotency should look like". He thinks that there are more states of pluripotency to be found, and his group will be looking for them in its hard drives. "It's a conceptually important thing, it opens up a big door," he says.

All these studies are adding fuel to a central debate in the reprogramming community: does the process have an inherently random and unpredictable element to it? Until recently, there was a general consensus that this was

David Cyranoski *reports for* Nature *from Shanghai.*

1. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
2. Hussein, S. M. *et al. Nature* **516,** 198–206 (2014).
3. Tonge, P. D. *et al. Nature* **516,** 192–197 (2014).
4. Clancy, J. L. *et al. Nature Commun.* http://dx.doi.org/10.1038/ncomms6522 (2014).
5. Benevento, M. *et al. Nature Commun.* http://dx.doi.org/10.1038/ncomms6613 (2014).
6. Lee, D.-S. *et al. Nature Commun.* http://dx.doi.org/10.1038/ncomms6619 (2014).
7. Okita, K., Ichisaka, T. & Yamanaka, S. *Nature* **448,** 313–317 (2007).
8. Soufi, A., Donahue, G. & Zaret, K. S. *Cell* **151,** 994–1004 (2012).
9. Buganim, Y., Faddah, D. A. & Jaenisch, R. *Nature Rev. Genet.* **14,** 427–439 (2013).
10. Koche, R. P. *et al. Cell Stem Cell* **8,** 96–105 (2011).
11. Buganim, Y. *et al. Cell* **150,** 1209–1222 (2012).
12. Di Stefano, B. *et al. Nature* **506,** 235–239 (2014).

COMMENT

EYE OF SCIENCE/SPL
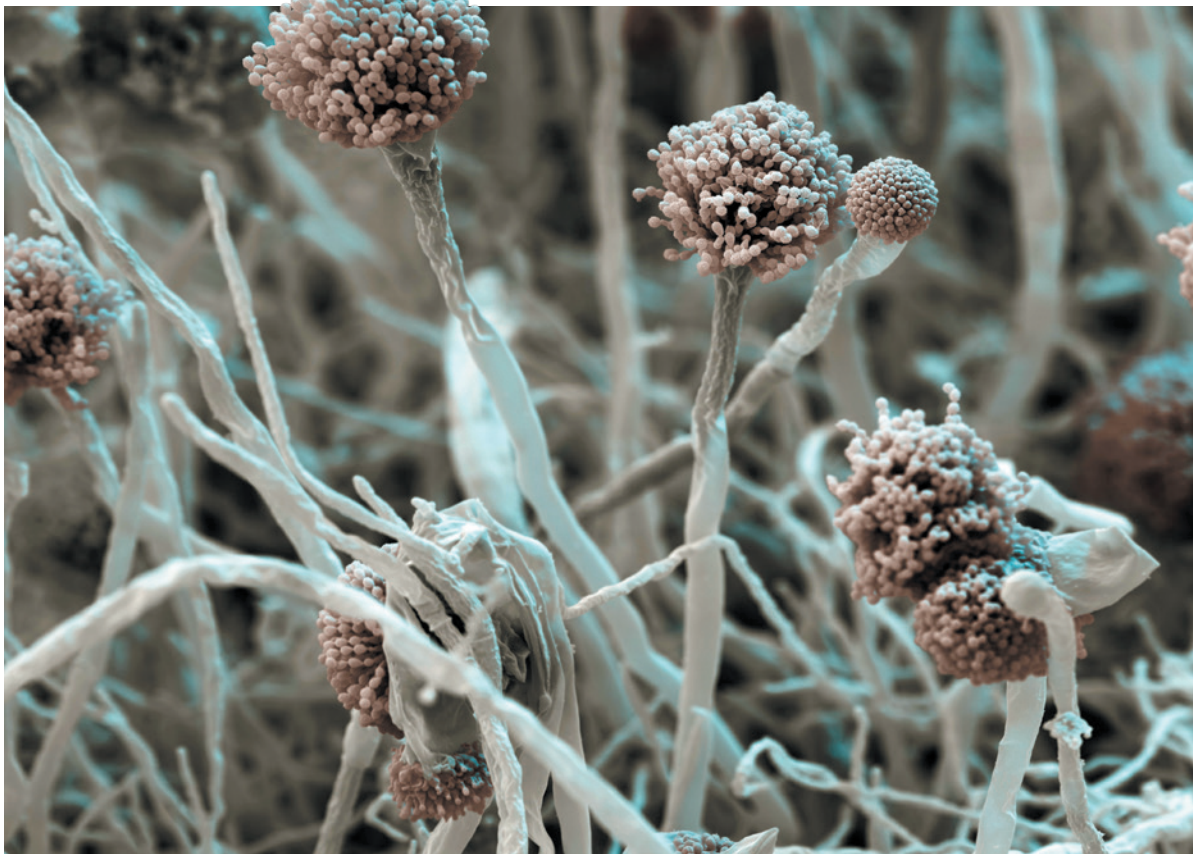
The often harmless fungus *Aspergillus fumigatus* can cause severe pulmonary disease in people with leukaemia.

Ditch the term pathogen

Disease is as much about the host as it is the infectious agent — the focus on microbes is hindering research into treatments, say **Arturo Casadevall** and **Liise-anne Pirofski**.

The term pathogen started to be used in the late 1880s to mean a microbe that can cause disease. Ever since, scientists have been searching for properties in bacteria, fungi, viruses and parasites that account for their ability to make us ill. Some seminal discoveries have resulted — such as the roles of various bacterial and fungal toxins in disease. Indeed, our oldest and most reliable vaccines, such as those for diphtheria and tetanus, work by prompting the body to produce antibodies that neutralize bacterial toxins.

Yet a microbe cannot cause disease without a host. What actually kills people with diphtheria, for example, is the strong inflammatory response that the diphtheria toxin triggers, including a thick grey coating on the throat that can obstruct breathing. Likewise, it is the massive activation of white blood cells triggered by certain strains of *Staphylococcus* and *Streptococcus* bacteria that can lead to toxic-shock syndrome.

Disease is one of several possible outcomes of an interaction between a host and a microbe. It sounds obvious spelled out in this way. But the issue here is more than just semantics: the use of the term pathogen sustains an unhelpful focus among researchers and clinicians on microbes that could be hindering the discovery of treatments. In the current Ebola epidemic in West Africa, for instance, much attention has been focused on the ill and the dead, even though crucial clues to curbing the outbreak may ▶

▶ be found in those who remain healthy despite being exposed to the virus.

Instead of focusing on what microbes do or do not do[1], researchers should ask whether an interaction between a host and a microbe damages the host, and if so, how. This approach will require different tools and potentially more alliances between microbiologists and immunologists.

CONTEXT IS EVERYTHING

In the decades after the word pathogen was coined, it became clear that many 'non-pathogens' can be harmful in some people. Until the 1950s, for example, coagulase-negative staphylococci (part of the normal flora of human skin) and *Candida albicans* (usually present in the vagina, mouth and gut, and on the skin) were rarely associated with disease. Infections caused by these microbes then became common with the use of intravenous catheters, which open a channel between the skin and the blood, and treatments that suppress immunity, such as chemotherapy.

This prompted microbiologists to use qualifiers, mostly from the 1960s onwards, to define microbes according to their state in the host organism. For instance, 'commensal' was used to describe microbes that live on or in hosts without causing harm, such as *Escherichia coli*, one of the many species present in the human gut; 'colonizer' referred to organisms commonly found in the human body but able to cause disease, such as *Staphylococcus*; 'saprophyte' described organisms associated with dead plant material, including the fungus *Aspergillus fumigatus*.

But even these qualifiers proved inadequate. Microbes and hosts are variable and unpredictable. For instance, *A. fumigatus* can cause severe pulmonary disease in people with leukaemia; some strains of *E. coli* can cause diarrhoea and vomiting, and in one out of three people, *Staphylococcus aureus* behaves more like a commensal, inhabiting nasal cavities without causing harm.

During the 1970s, biologists began to try to identify microbial genes that confer pathogenicity. Researchers deleted or inactivated genes in search of those encoding 'virulence factors', molecules thought to enable a microbe to invade and inhabit a host and cause disease. This hunt for microbial genes or mutations associated with disease continues to this day. For example, researchers are applying genomics to try to discern signatures of virulence among *S. aureus*, *Haemophilus influenzae* and *Enterococcus faecium* strains, to name a few[2–4].

"Much of the research on infectious diseases continues to be dominated by reductionist approaches."

The approach has worked extremely well for some bacteria. For example, knocking out the toxin and capsule genes of *Bacillus anthracis* rendered the bacterium less virulent, and so suitable for use in a vaccine against anthrax. It has been less successful with other microbes, such as types of fungus. More than two decades of research have been devoted to trying to find microbial factors that enable *C. albicans* and *A. fumigatus* to cause disease. In neither case does a single classical virulence factor seem to have a big effect on pathogenicity.

VACCINE CHALLENGES

Work on vaccines has provided further indications of there being flaws in the idea that discrete factors, akin to toxins, enable all microbes to cause disease.

Most vaccine research has focused on identifying and neutralizing microbes' virulence factors. In numerous cases, this tactic has paid off. The vaccines for tetanus and diphtheria work on this basis, and have eliminated two major killers from the Western world. Similarly, a vaccine that makes the polysaccharide capsule of bacteria vulnerable to attack from white blood cells by prompting lymphocytes to produce antibodies against it has virtually eradicated *H. influenzae* type B, a major cause of meningitis before the 1980s. Since 2000, similar vaccines have markedly reduced the incidence of disease caused by *Streptococcus pneumoniae*.

Yet at least for *S. pneumoniae*, the idea that antibodies prevent disease solely by promoting uptake and killing of the microbe by immune cells called phagocytes is too simplistic. The mere presence of antibodies to *S. pneumoniae* in someone's blood, for instance, does not reliably indicate that the person will be protected from pneumonia. What is more, many of the ongoing attempts to develop new vaccines by identifying and targeting virulence factors have so far proved fruitless. Despite decades of searching, no classical virulence factor suitable for vaccine development has been identified for the tuberculosis bacillus or malaria parasite.

In some cases, efforts aiming to neutralize virulence factors may even have uncovered ways to exacerbate disease. Pulmonary tuberculosis occurs in less than 10% of people infected with *Mycobacterium tuberculosis*. In these people, an over-exuberant inflammatory response destroys lung tissue. Thus, vaccines against tuberculosis that are designed to enhance the immune response might not work.

This could explain why in the 1890s, when microbiologist Robert Koch injected people who had tuberculosis with an extract that he had produced from culturing the bacteria in the laboratory, many of them died. It could also explain why certain

vaccines produced in the past century, for instance for the respiratory syncytial virus, failed to prevent disease.

CHANGING DYNAMICS

The term pathogen is unlikely to go away. But those who study infectious diseases need to own up to its limitations.

Researchers probing the human microbiome (the community of microorganisms that live in and on our bodies) using genomics are being forced to recognize that myriad factors and interactions shape its composition. It varies in different people, at different times in development and in association with disease.

Yet much of the research on infectious diseases continues to be dominated by reductionist approaches; one variable is altered while all others are assumed to hold constant. Microbiologists tend to view the microbe as the key variable in disease and treat the host as a constant. Immunologists generally see the microbe as a constant and the host response as the variable (for instance, immunologists frequently inject microbes into normal and genetically manipulated laboratory animals, to assess the factors that shape the host response)[5]. These two groups go to different conferences, read and publish in different journals, and receive funds from different granting panels.

What is needed is the simultaneous analysis of microbial and host variables using new analytical tools. Damage to the host is a measurable parameter that can result from the microbe, the host's response, or both, and as such, it shifts the focus onto the host–microbe interaction[6].

New tools are needed to measure the spectrum of inflammatory, biochemical and other forms of damage resulting from the interaction between hosts and microbes. The discovery and development of these tools must be driven by new sessions at conferences, special issues of journals and dedicated funding streams. We think that such a shift in approach would uncover all sorts of possibilities for preventing infectious diseases. ■

Arturo Casadevall *and* **Liise-anne Pirofski** *are professors of medicine, microbiology and immunology at the Albert Einstein College of Medicine, Yeshiva University, New York, USA.*
e-mails: arturo.casadevall@einstein.yu.edu; l.pirofski@einstein.yu.edu

1. Pirofski, L. A. & Casadevall, A. *BMC Biol.* **10,** 6 (2012).
2. Howden, B. P. *et al. mBio* **4,** e00412-13 (2013).
3. De Chiara, M. *et al. Proc. Natl Acad. Sci. USA* **111,** 5439–5444 (2014).
4. Young, B. C. *et al. Proc. Natl Acad. Sci. USA* **109,** 4550–4555 (2012).
5. Biron, C. A. & Casadevall, A. *mBio* **1,** e00260-10 (2010).
6. Casadevall, A. & Pirofski, L. A. *Nature Microbiol. Rev.* **1,** 17–24 (2003).

The protagonists of Goethe's science novel compare their changing attractions to chemical bonding,

IN RETROSPECT
Elective Affinities

Matthew Bell reassesses German polymath Goethe's haunting 'chemical romance'.

In 1809, German national poet and polymath Johann Wolfgang von Goethe provoked outrage with his novel *Elective Affinities* (*Die Wahlverwandtschaften*). Many readers were horrified by its almost playful treatment of adultery, which still carries a charge. But what makes it worthy of reappraisal is how it puts science at the centre of human concerns — and humans at the centre of science. Goethe emphasized the primacy of human perception in understanding nature as a holistic entity, in contrast to the quantitative methods and mechanical Universe of the Enlightenment era.

Today, the basic phenomena of biology and physics have been described, and the complexity of the Universe and the human brain are testing reductionism. Multiple perspectives and big data are needed to crack the challenges of energy, food and population. People are at the centre of research — where Goethe felt they must be. A reappraisal of Goethe's science reminds us how key advances emerged outside the mechanical tradition of Descartes and Newton.

Goethe began his scientific studies in the late eighteenth century. Biology was largely an observational science, dependent on skill and patience; Goethe's notebooks are full of meticulous descriptions of plants, mammals and insects. He raised theoretical questions about the nature and origins of life and the development of species, for which the mechanical model had no plausible answers. Resisting the urge to jump to conclusions, Goethe focused on amassing data and seeking patterns. He painstakingly documented the relations between parts of organisms and the similarities between species. From observations of hundreds of skeletons, he developed an influential model of mammal anatomy. In seeking a principle underlying organic nature, he is seen as a founder of modern biology; Charles Darwin mentioned him in the 'historical sketch' in the 1860 second edition of *On the Origin of Species*.

Elective Affinities weaves together many of these strands. The title comes from the work of Swedish chemist Torbern Bergman, who devised the eighteenth century's most accurate chart of what was likely to bind with what — a forerunner of the periodic table. Bergman's theory of 'elective affinities' seems to describe the shifting relationships of the protagonists, Eduard, Charlotte, Ottilie and the Captain. In this sense, the novel can be read as an exercise in reductionism: like elements, the characters seem to have no choice but to make new bonds when a reagent is introduced. Even their names reinforce this. Both Eduard and the Captain were christened Otto, so the repetition of 'ott' in the names of the characters emerges as a sign of affinity.

Early on, the Captain and the married Eduard and Charlotte discuss Bergman's theory and its application to relationships. They are not merely objects in Goethe's experiment: they consciously experiment on themselves. Goethe believed that experiments and the experimenter are one — and that a human is the most precise apparatus.

The novel is set on Eduard and Charlotte's estate. Eduard persuades his wife to invite her orphaned niece Ottilie and his friend the Captain to stay with them as an 'experiment' (*Versuch*), establishing empiricism as a metaphor for human relations. The Captain and Eduard begin to improve the estate, leaving Eduard less time with Charlotte; they compare

Johann Wolfgang von Goethe.

this change in their bonds to a reaction. Eduard jokes that Ottilie will form a 'compound' with Charlotte, but himself comes to see an affinity with her. Charlotte and the Captain are drawn to each other. When Eduard and Charlotte make love, their minds are occupied with thoughts of these others. It ends tragically. Charlotte gives birth to a son (another Otto), whom Ottilie accidentally drowns; Ottilie starves to death, followed by Eduard.

Beyond its reading as a human analogue of chemical reactions, the novel is infused with aspects of Goethe's science expressed in his quixotic *Theory of Colour*, published the following year. Goethe believed that Newton's experiments with prisms were flawed. He contended that white light was the fundamental phenomenon, and that colours were produced by interactions between light and darkness, perceivable by the naked eye — incorrect, but an accurate record of how we perceive colour. However, it is Goethe's argument that Newton valued representation of phenomena in symbols over the phenomena themselves that has resonance in *Elective Affinities*.

Goethe's novel can be seen as an attempt to show the consequences of the urge to abstraction. The narcissistic Eduard interprets isolated phenomena, such as headaches on the right side of his head and the left of Ottilie's, as symbols of affinity, and hypothesizes from that. Goethe believed that scientists should critically observe a broad spectrum of phenomena before theorizing. Eduard, driver of the tragic plot, is Goethe's personification of the flaws that he found in the science of his day.

Although *Elective Affinities* scandalized nineteenth-century readers, its theme and penetration sparked a cult following among writers. George Eliot — whose unmarried relationship with Goethe scholar George Henry Lewes was itself a scandal — admired the novel, and it may have influenced her harrowing *The Mill on the Floss* (1860). Characters and plots in Ford Madox Ford's *The Good Soldier* (1915) echo it, and protagonists of John Banville's *The Newton Letter* (1982) are named Edward, Charlotte and Ottilie.

Goethe called for a "gentle empiricism", believing that advanced human development (*Bildung*) was essential to the perception of nature's wondrous realities. *Elective Affinities*, by questioning the fruits of reductionism, challenges us to recall that no observer can ever be impartial. ∎

Matthew Bell *is professor of German and comparative literature at King's College London. His books include* Melancholia: The Western Malady. *He has edited a forthcoming translation of Goethe's works. e-mail: matthew.bell@kcl.ac.uk*

Calvin Bridges experimented on fruit flies to make fundamental discoveries in genetics.

GENETICS

Genius on the fly

Ewen Callaway reviews a biopic of Calvin Bridges, the wild-living, wild-haired genetics pioneer.

Calvin Bridges is best known for three things: his pioneering work on genetics in the early twentieth century, his womanizing and his gravity-defying mop of hair. *The Fly Room*, a biopic told through the eyes of his daughter Betsey, also shows the scientist as a sometimes dedicated, often distracted father who struggled to balance intellectual curiosity with family obligations.

The Fly Room came from a chance encounter between geneticist-turned-filmmaker Alexis Gambis and Betsey, now in her nineties. It is bookended by interviews with her, but focuses on a period in the 1920s, when ten-year-old Betsey visited her father's workplace: the famed Fly Room at Columbia University in New York City. The film was partly crowdfunded through Kickstarter. Researchers including neuroscientists Joseph LeDoux and Stuart Firestein have supporting roles.

Bridges, portrayed by a wild-haired Haskell King, was a star disciple of evolutionary biologist Thomas Hunt Morgan (played by Firestein). Under Morgan's leadership, Bridges and a cadre of Young Turks characterized mutant fruit flies — most famously, white-eyed varieties — to map the locations of genes and to understand how they are transmitted. Bridges' work established that trait-determining genes are carried by chromosomes that parents pass to their offspring. He also worked out how chromosomes — X and Y — determine the sex of fruit flies.

"It was not unusual for six of us to carry on in this small room," Morgan remembered in an obituary of Bridges, who died in 1938 from

The Fly Room
WRITER/DIRECTOR:
ALEXIS GAMBIS
Imaginal Disc: 2014.

syphilis. To feed the flies, near-rotting bananas were a constant presence.

Those bananas dangle from the ceiling in the film's fictionalized Fly Room, where Bridges and other prominent figures in genetics sort dead flies and trade rude witticisms. Betsey crashes this world after her mother, Gertrude, sends her to spend time with her father. None of the scientists knows what to make of the curious girl, who carries her box camera everywhere. Bridges is annoyed to have his sanctum disturbed. But he warms to Betsey and puts her to use in the Fly Room, counting and characterizing flies. He becomes so comfortable having his daughter around that he neglects to hide his after-hours philandering.

Much of the film unfolds in the Fly Room. The set designers have paid close attention to detail: for example, the microscopes are the binocular version that Bridges invented.

Bridges left his family; Morgan moved his lab to the California Institute of Technology and Bridges joined him. *The Fly Room* makes no attempt to provide an authoritative history, leaving many details to the epilogue. In an interview, 95-year-old Betsey says that she never wanted to be like her dad. An apt sentiment about a father who was flawed — but who laid the groundwork for the modern science of heredity. ∎

Ewen Callaway *writes for* Nature *from London.*

Correspondence

Manage military land for the environment

A refocus on managing military training grounds for their value to the environment as well as to the armed forces would drastically increase the global terrestrial 'protected area' at minimal cost (see J. E. M. Watson et al. Nature 515, 67–73; 2014).

We estimate that training areas total at least 50 million hectares, with the actual figure probably closer to 300 million hectares (R. Zentelis and D. Lindenmayer Conserv. Lett., in the press). These areas encompass all major global ecosystems, including those poorly represented within formal reserve systems. In the Western world, at least, their management is already funded through military expenditure.

Many examples highlight the value of such areas. They support the majority of Germany's wolf packs, and in Australia they contain some of the best remaining threatened coastal heathland. Regardless of one's view of the military, the armed forces manage a huge area of land that, until now, has not been recognized as an important funded conservation resource.

Rick Zentelis, David Lindenmayer Australian National University, Canberra, Australia.
rick.zentelis@anu.edu.au

Europe is failing young researchers

We are young European researchers and participants in science-policy initiatives who feel strongly that the European Research Area (ERA) faces many challenges.

The absence of a fully inclusive and self-sufficient ERA still affects research institutions locally. Regional funding remains too sparse and fragmented. As well as a dearth of sustainable career opportunities, there is widespread cronyism, and many administrative and research structures are obsolete.

We need more transparency and objectivity in funding, promotions and hiring practices. Such reforms would cost relatively little and might even make some funding cuts unnecessary.

The responsibility for improvement lies not only with the European governing bodies, but also with member states and regions. These are issues on which the undersigned all agree — we are members of the COST Sci-Generation Network, the Young Academy of Europe, the Global Young Academy and EURAXESS Voice of the Researchers.

Thomas Schäfer* Polymat, University of the Basque Country, Donostia-San Sebastián; and Ikerbasque, Bilbao, Spain.
thomas.schafer@ehu.es
*On behalf of 15 correspondents (see go.nature.com/ab6jtb for full list).

Biodiversity reports need author rules

Two representatives from the agrochemical industry are among 40 authors of a fast-track assessment of pollinators by the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES; see go.nature.com/q8lll2). In our view, to support the credibility of assessment results, the IPBES needs a policy requiring authors to declare all funding sources, positions held and other potential conflicts of interest.

It is unclear how the IPBES deals with conflicts of interest. Their second plenary meeting last December postponed a decision on the matter. Authors are nominated by IPBES member states and other stakeholders to "reflect the range of scientific, technical and socio-economic views and expertise; geographical representation … ; the diversity of knowledge systems … ; and gender balance". But the IPBES has no explicit rules for nomination or selection.

IPBES assessments could lead to far-reaching policy interventions, with financial implications for industry sectors (for example, in mining after assessment of land degradation and restoration, or for transport after invasive-species assessment). Given the role of agrochemicals in pollinator decline (J. van der Sluijs et al. Environ. Sci. Pollut. Res. http://doi.org/xcx; 2014), it is our view that scientists funded by such corporations should not be lead authors or coordinating lead authors on such assessments.

We also suggest that the IPBES publishes the names of all nominated authors, along with their nominators and justification for their appointment.

Axel Hochkirch Trier University, Germany.
Philip J. K. McGowan Newcastle University, UK.
Jeroen van der Sluijs University of Bergen, Norway.
hochkirch@uni-trier.de

Engaged cohort good for science

As staff at the UK Avon Longitudinal Study of Parents and Children (ALSPAC), we agree that participant involvement is crucial to the design of cohort studies (P. Lucas et al. Nature 514, 567; 2014). We work with an advisory panel composed of a large and representative selection of original cohort participants.

The panel provides regular, thoughtful feedback and advice to ALSPAC researchers about data-collection exercises. It comments on proposals, the appropriateness of questions, communications materials and channels, research findings and the burden on participants. This helps to improve our study and makes the broader cohort more likely to engage in our research.

We also host focus groups and online discussion forums with all segments of our cohort — mothers, fathers, siblings and young parents — and use Facebook and Twitter. ALSPAC is cited as an example of best social-media practice in guidelines from the UK National Institute for Health Research (see go.nature.com/txsxma).

We look for new ways to hear participants' views, on topics from our newsletters to a 2012 events programme (see Nature 484, 155–158; 2012). Devised by participants to mark their 21st birthdays, this included a science festival, a conference, parties for study children and parents, and a commemorative book.

Katarzyna Kordas, Dara O'Hare, Makaela Jacobs-Pearson University of Bristol, UK.
kasia.kordas@bristol.ac.uk

Several fields still need primates

Eliminating the use of non-human primates in certain fields (see P. Bateson and C. I. Ragan Nature 514, 567; 2014) has no bearing on their utility in neuropsychiatry and neurology.

The use of these animals, including genetically modified marmosets, is in our view essential for fundamental research into mental-health disorders. Similarities in the structure of higher-order cortical brain regions — which are dysregulated in disorders such as depression and schizophrenia — enable the most accurate and relevant mapping of the primate brain's functional organization.

A prominent example is the mapping of neural pathways in the rhesus monkey, which led to the discovery that deep brain stimulation can be an effective treatment for Parkinson's disease (see go.nature.com/28spre).

The US National Institute of Mental Health has recognized that such fundamental research should be applied to the understanding and treatment of neuropsychiatric disorders (Research Domain Criteria; see go.nature.com/or4keu), to identify discrete psychological deficits associated with specific neural pathways.

Angela Roberts, Trevor Robbins University of Cambridge, UK.
acr4@cam.ac.uk

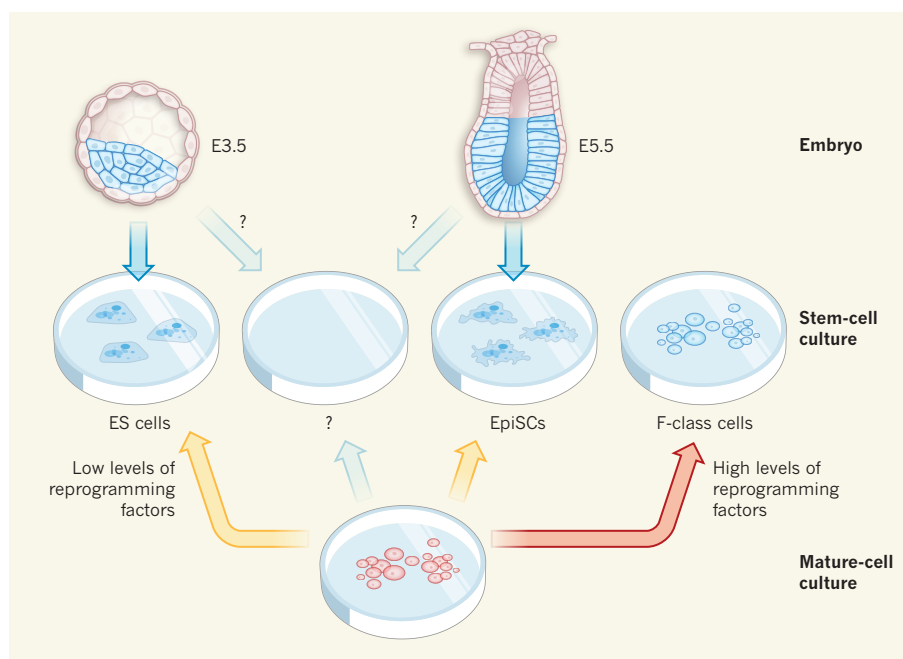**STEM CELLS**

A designer's guide to pluripotency

Pluripotent stem cells, which give rise to almost all cell types, can be engineered from mature cells. A thorough analysis of the process has led to the characterization of a new type of pluripotent cell. SEE ARTICLES P.192 & P.198

JUN WU & JUAN CARLOS IZPISUA BELMONTE

Pluripotency, defined as the ability of a cell to generate all cell types in the adult organism, is a transient feature of early embryonic development. Two distinct pluripotent cell types can be isolated from embryos and cultured *in vitro*[1–4] — naive cells, called embryonic stem cells, and those primed for differentiation, epiblast stem cells. Furthermore, a defined cocktail of transcription factors, called reprogramming factors, can reinstate pluripotency when introduced into mature cells, producing induced pluripotent stem cells (iPSCs)[5–7]. In addition to known pluripotent cell types[5,8], iPSC generation yields a spectrum of distinct cell types, hinting at the existence of uncharacterized pluripotent states. A collection of five manuscripts (two in this issue[9,10] and three in *Nature Communications*[11–13]), now uncover and characterize an alternative pluripotent outcome of iPSC reprogramming: F-class cells (Fig. 1).

These five manuscripts are part of an international collaboration called Project Grandiose, in which the researchers set out to reanalyse the process of iPSC reprogramming from an unbiased perspective. They reasoned that, by extensively documenting the molecular and cellular transitions occurring at each stage of the process, they could provide both the first thorough roadmap for iPSC reprogramming, and an explanation for the emergence during reprogramming of undefined pluripotent cell types, which have been mostly overlooked by previous studies.

In the first paper, Tonge *et al.*[9] (page 192) identify F-class cells — named because of their unusual, fuzzy-looking colony morphology — as a pluripotent cell type distinct from embryonic stem cells (ES cells) and epiblast stem cells. Maintenance of F-class cells depends on continuing high expression of reprogramming factors. In conventional reprogramming methods, the expression of introduced genes (transgenes) is silenced by factors that are expressed in the host cells once pluripotency is achieved, and thus F-class cells could not have been identified in those assays. The researchers' use of a host-factor-independent reprogramming method bypasses transgene silencing and thereby

Figure 1 | Different flavours of pluripotency. Two distinct types of pluripotent stem cell have been captured from early mouse embryos for culture *in vitro* — embryonic stem cells (ES cells) from embryos at three-and-a-half days old (E3.5), and epiblast stem cells (EpiSCs) from embryos at E5.5. The pluripotent-cell populations in each embryo are shown in blue. These two cell types can also be induced from mature cells through cellular reprogramming using low levels of reprogramming factors. Five papers[9–13] from Project Grandiose investigate the molecular details of cellular reprogramming, and uncover a new type of pluripotent cell, dubbed F-class, which depends on sustained, high-level expression of reprogramming factors. This discovery hints at the potential that other, unidentified pluripotent states exist (marked with a question mark), and might either be generated by engineering or be present in the early embryo.

allows sustained high-level expression of reprogramming factors[14].

Tonge and colleagues report that the fuzzy morphology of F-class cells arises from their low adhesiveness, which, along with their fast proliferation, makes these cells more amenable to large-scale production than ES cells. This is a desirable feature for cell-based therapies, which demand large quantities of specific cell types. For example, pancreatic β-cells, which store and release insulin, can be derived from pluripotent cells and might be used to treat people with diabetes[15]. However, F-class cells' dependence on transgenes could be problematic for their safe clinical application, because mutations arising from either improper transgene insertion into the genome

or incomplete inactivation of reprogramming factors when the cells begin differentiation might ultimately lead to tumour formation.

One solution might be to stabilize the F-state independent of transgenes, using small molecules. This strategy has been successful for stabilizing naive-like human pluripotent stem cells[16,17]. Tonge *et al.* show that ES-like cells convert to the F-state following forced expression of reprogramming factors. Conversely, F-class cells can be converted to an ES-cell-like state using small molecules that inhibit the activity of a class of enzymes called histone deacetylases, which modulate gene expression by removing acetyl molecules from the histone proteins around which DNA is packaged. Such interconvertibility may lead to insights

into how pluripotency is stabilized in distinct cellular contexts.

In the second paper, Hussein *et al.*[10] (page 198) define the different molecular routes to pluripotency by performing the most detailed analysis of reprogramming so far. Among other findings, the authors uncover key determinants for the emergence of ES-cell-like or F-class states. Emergence of the F-class state relies on repression of genes that are expressed in ES cells. This is achieved through a molecular modification associated with gene repression — the attachment of three methyl molecules to an amino-acid residue, lysine 27, of histone H3 proteins. By contrast, the loss of the DNA methylation marks inherited from mature cells is necessary for cells to take on an ES-cell-like state, but some of these marks are retained in F-class cells.

The remaining three studies complement Hussein and colleagues' work by providing descriptive, in-depth analyses of the changes in molecular pathways en route to pluripotency, generating large data sets that are freely available at www.stemformatics.org. Lee *et al.*[11] interrogate the epigenetic changes (those modifications to the genome that affect gene expression without altering DNA sequence) that occur during the transition to pluripotency. They conclude that DNA methylation has a crucial role in iPSC reprogramming and acts as an epigenetic switch between F-class and ES-cell-like states. Clancy and colleagues[12] delineate the dynamic changes in small RNAs — post-transcriptional regulators of gene expression — during iPSC reprogramming, and find that a distinct group of

microRNAs supports the F-class pluripotency program. Finally, Benevento *et al.*[13] show that reorganization of protein expression occurs in two defined waves during cellular reprogramming. The authors show that patterns of protein expression differ between ES-cell-like and F-class states.

These five manuscripts mark the first steps towards understanding F-class pluripotency and thus towards making the most of their clinical potential. The molecular mechanisms underpinning the F-state warrant further investigation, as do the metabolic cues that contribute to sustaining F-class cells, because different pluripotent stem cells probably have distinct metabolic requirements[18]. Remaining questions include whether human F-class cells can be generated through cellular reprogramming, and if functional differentiated cells can be obtained from F-class cells.

In embracing the inherent artificiality of iPSC reprogramming, Project Grandiose has opened up the field to fresh avenues of research. This work shows that a third pluripotent state can be engineered *in vitro*, and it may be that there are other pluripotent endpoints of reprogramming (Fig. 1). Moreover, there may be other pluripotent states in the developing embryo. If there are, it would be interesting to determine whether such states could be captured and cultured *in vitro*. To investigate these avenues, an unbiased approach, such as that taken by Tonge *et al.*, will probably prevail.

Looking ahead, customized stem cells designed for specific applications — such as large-scale expansion, or fast, synchronized differentiation — may soon become a reality.

The existence of alternative pluripotent states adds another dimension to the potential of pluripotent stem cells in regenerative medicine. The results of Project Grandiose call for future work that catalogues myriad molecularly and functionally distinct pluripotent stem cells to harness their full potential. ■

Jun Wu *and* Juan Carlos Izpisua Belmonte *are in the Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA.*
e-mail: belmonte@salk.edu

1. Evans, M. J. & Kaufman, M. H. *Nature* **292,** 154–156 (1981).
2. Martin, G. R. *Proc. Natl Acad. Sci. USA* **78,** 7634–7638 (1981).
3. Brons, I. G. *et al. Nature* **448,** 191–195 (2007).
4. Tesar, P. J. *et al. Nature* **448,** 196–199 (2007).
5. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
6. Shu, J. *et al. Cell* **153,** 963–975 (2013).
7. Montserrat, N. *et al. Cell Stem Cell* **13,** 351–350 (2013).
8. Han, D. W. *et al. Nature Cell Biol.* **13,** 66–71 (2010).
9. Tonge, P. D. *et al. Nature* **516,** 192–197 (2014).
10. Hussein, S. M. I. *et al. Nature* **516,** 198–206 (2014).
11. Lee, D. S. *et al. Nature Commun.* **5,** 5619; http://dx.doi.org/10.1038/ncomms6619 (2014).
12. Clancy, J. L. *et al. Nature Commun.* **5,** 5522; http://dx.doi.org/10.1038/ncomms6522 (2014).
13. Benevento, M. *et al. Nature Commun.* **5,** 5613; http://dx.doi.org/10.1038/ncomms6613 (2014).
14. Woltjen, K. *et al. Nature* **458,** 766–770 (2009).
15. Pagliuca, F. W. *et al. Cell* **159,** 428–438 (2014).
16. Gafni, O. et al. *Nature* **504,** 282–286 (2013).
17. Theunissen, T. W. *et al. Cell Stem Cell* **15,** 471–487 (2014).
18. Zhou, W. *et al. EMBO J.* **31,** 2103–2116 (2012).

Breakthrough for protons

The atomically thin material called graphene is impermeable to atoms as small as helium. The finding that protons can pass through it might enable new kinds of membrane to be developed and aid research into fuel cells. SEE LETTER P.227

ROHIT N. KARNIK

The two-dimensional material graphene is often depicted as a hexagonal mesh of carbon atoms, with plenty of space between its atoms. But in reality, the finite size of the carbon atoms leaves little room for anything to slip through. In 2008, a classic experiment[1] revealed that pristine graphene is impermeable to helium and other gases at room temperature, making it the thinnest barrier known to science. The results logically extend to other two-dimensional materials,

including hexagonal boron nitride (hBN) and molybdenum disulphide (MoS_2). By contrast, in a paper published in this issue (page 227), Hu *et al.*[2] present the unexpected finding that graphene and hBN — but not MoS_2 — are excellent conductors of protons across their two-dimensional structure.

The authors measured the electric current across micrometre-sized flakes of graphene, hBN or MoS_2 sandwiched between two layers of a polymer that conducts protons when hydrated (that is, in the presence of water). In the absence of other charge carriers, the

measured current is a direct indicator of proton transport. Hu and colleagues detected substantial current across graphene, and an even higher current across hBN, but no current across MoS_2 — indicating that graphene and hBN conduct protons, but MoS_2 does not.

Any contribution to the proton conductivity from defects in the two-dimensional materials can be ruled out, because the results were remarkably repeatable across different experiments and because the researchers carefully characterized the materials. The same conductivities were obtained when aqueous hydrochloric acid — a source of protons — was placed on either side of the materials, showing that the proton conductivity was a general effect and was not limited to the experiment with polymer layers. Hu and co-workers went on to show that bilayers and trilayers of hBN conduct protons, albeit with reduced conductivity compared with monolayers. However, in the case of graphene, even one extra layer entirely obliterates proton conductivity, so that bilayer graphene is essentially impermeable.

The observed proton conductivities — or lack thereof — can be explained by the electron-density distribution in the

into how pluripotency is stabilized in distinct cellular contexts.

In the second paper, Hussein *et al.*[10] (page 198) define the different molecular routes to pluripotency by performing the most detailed analysis of reprogramming so far. Among other findings, the authors uncover key determinants for the emergence of ES-cell-like or F-class states. Emergence of the F-class state relies on repression of genes that are expressed in ES cells. This is achieved through a molecular modification associated with gene repression — the attachment of three methyl molecules to an amino-acid residue, lysine 27, of histone H3 proteins. By contrast, the loss of the DNA methylation marks inherited from mature cells is necessary for cells to take on an ES-cell-like state, but some of these marks are retained in F-class cells.

The remaining three studies complement Hussein and colleagues' work by providing descriptive, in-depth analyses of the changes in molecular pathways en route to pluripotency, generating large data sets that are freely available at www.stemformatics.org. Lee *et al.*[11] interrogate the epigenetic changes (those modifications to the genome that affect gene expression without altering DNA sequence) that occur during the transition to pluripotency. They conclude that DNA methylation has a crucial role in iPSC reprogramming and acts as an epigenetic switch between F-class and ES-cell-like states. Clancy and colleagues[12] delineate the dynamic changes in small RNAs — post-transcriptional regulators of gene expression — during iPSC reprogramming, and find that a distinct group of

microRNAs supports the F-class pluripotency program. Finally, Benevento *et al.*[13] show that reorganization of protein expression occurs in two defined waves during cellular reprogramming. The authors show that patterns of protein expression differ between ES-cell-like and F-class states.

These five manuscripts mark the first steps towards understanding F-class pluripotency and thus towards making the most of their clinical potential. The molecular mechanisms underpinning the F-state warrant further investigation, as do the metabolic cues that contribute to sustaining F-class cells, because different pluripotent stem cells probably have distinct metabolic requirements[18]. Remaining questions include whether human F-class cells can be generated through cellular reprogramming, and if functional differentiated cells can be obtained from F-class cells.

In embracing the inherent artificiality of iPSC reprogramming, Project Grandiose has opened up the field to fresh avenues of research. This work shows that a third pluripotent state can be engineered *in vitro*, and it may be that there are other pluripotent endpoints of reprogramming (Fig. 1). Moreover, there may be other pluripotent states in the developing embryo. If there are, it would be interesting to determine whether such states could be captured and cultured *in vitro*. To investigate these avenues, an unbiased approach, such as that taken by Tonge *et al.*, will probably prevail.

Looking ahead, customized stem cells designed for specific applications — such as large-scale expansion, or fast, synchronized differentiation — may soon become a reality.

The existence of alternative pluripotent states adds another dimension to the potential of pluripotent stem cells in regenerative medicine. The results of Project Grandiose call for future work that catalogues myriad molecularly and functionally distinct pluripotent stem cells to harness their full potential. ■

Jun Wu *and* **Juan Carlos Izpisua Belmonte** *are in the Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA.*
e-mail: belmonte@salk.edu

1. Evans, M. J. & Kaufman, M. H. *Nature* **292,** 154–156 (1981).
2. Martin, G. R. *Proc. Natl Acad. Sci. USA* **78,** 7634–7638 (1981).
3. Brons, I. G. *et al. Nature* **448,** 191–195 (2007).
4. Tesar, P. J. *et al. Nature* **448,** 196–199 (2007).
5. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
6. Shu, J. *et al. Cell* **153,** 963–975 (2013).
7. Montserrat, N. *et al. Cell Stem Cell* **13,** 351–350 (2013).
8. Han, D. W. *et al. Nature Cell Biol*. **13,** 66–71 (2010).
9. Tonge, P. D. *et al. Nature* **516,** 192–197 (2014).
10. Hussein, S. M. I. *et al. Nature* **516,** 198–206 (2014).
11. Lee, D. S. *et al. Nature Commun.* **5,** 5619; http://dx.doi.org/10.1038/ncomms6619 (2014).
12. Clancy, J. L. *et al. Nature Commun.* **5,** 5522; http://dx.doi.org/10.1038/ncomms6522 (2014).
13. Benevento, M. *et al. Nature Commun.* **5,** 5613; http://dx.doi.org/10.1038/ncomms6613 (2014).
14. Woltjen, K. *et al. Nature* **458,** 766–770 (2009).
15. Pagliuca, F. W. *et al. Cell* **159,** 428–438 (2014).
16. Gafni, O. *et al. Nature* **504,** 282–286 (2013).
17. Theunissen, T. W. *et al. Cell Stem Cell* **15,** 471–487 (2014).
18. Zhou, W. *et al. EMBO J.* **31,** 2103–2116 (2012).

MATERIALS SCIENCE

Breakthrough for protons

The atomically thin material called graphene is impermeable to atoms as small as helium. The finding that protons can pass through it might enable new kinds of membrane to be developed and aid research into fuel cells. SEE LETTER P.227

ROHIT N. KARNIK

The two-dimensional material graphene is often depicted as a hexagonal mesh of carbon atoms, with plenty of space between its atoms. But in reality, the finite size of the carbon atoms leaves little room for anything to slip through. In 2008, a classic experiment[1] revealed that pristine graphene is impermeable to helium and other gases at room temperature, making it the thinnest barrier known to science. The results logically extend to other two-dimensional materials,

including hexagonal boron nitride (hBN) and molybdenum disulphide (MoS$_2$). By contrast, in a paper published in this issue (page 227), Hu *et al.*[2] present the unexpected finding that graphene and hBN — but not MoS$_2$ — are excellent conductors of protons across their two-dimensional structure.

The authors measured the electric current across micrometre-sized flakes of graphene, hBN or MoS$_2$ sandwiched between two layers of a polymer that conducts protons when hydrated (that is, in the presence of water). In the absence of other charge carriers, the

measured current is a direct indicator of proton transport. Hu and colleagues detected substantial current across graphene, and an even higher current across hBN, but no current across MoS$_2$ — indicating that graphene and hBN conduct protons, but MoS$_2$ does not.

Any contribution to the proton conductivity from defects in the two-dimensional materials can be ruled out, because the results were remarkably repeatable across different experiments and because the researchers carefully characterized the materials. The same conductivities were obtained when aqueous hydrochloric acid — a source of protons — was placed on either side of the materials, showing that the proton conductivity was a general effect and was not limited to the experiment with polymer layers. Hu and co-workers went on to show that bilayers and trilayers of hBN conduct protons, albeit with reduced conductivity compared with monolayers. However, in the case of graphene, even one extra layer entirely obliterates proton conductivity, so that bilayer graphene is essentially impermeable.

The observed proton conductivities — or lack thereof — can be explained by the electron-density distribution in the
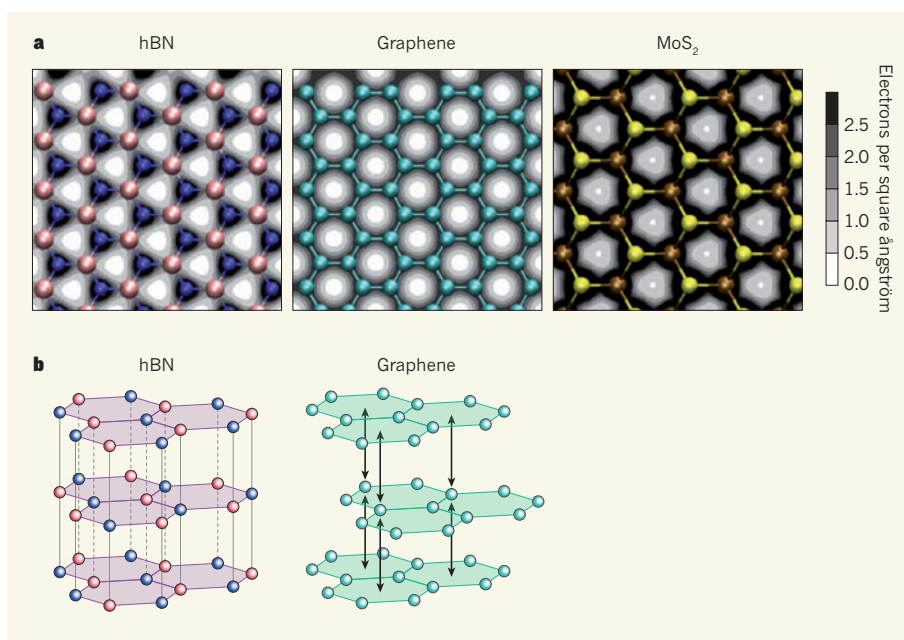
50 Years Ago

Goffman and Newill have directed attention to the analogy between the spreading of an infectious disease and the dissemination of information. We have recently examined the spreading of a rumour from the point of view of mathematical epidemiology … a mathematical model for the spreading of rumours can be constructed in a number of different ways … 'Reluctance to tell stale news' can be incorporated into the model.
From *Nature* 12 December 1964

100 Years Ago

In NATURE of December 3…there appeared a brief abstract of a paper communicated by Mr. Reginald A. Smith…on behalf of its author, Major E. R. Collins, D.S.O., now a wounded prisoner of war in Germany. This paper is not only an important contribution to our knowledge of the prehistoric stone implements of South Africa, but is evidence that a brave and capable soldier may, while helping to shape the history of his own time, give material assistance in unravelling the past history of the country through which he may be campaigning. Major Collins collected the material for his paper while engaged on trenching operations during the late Boer war … Major Collins made his collection of the stone industries of the ancient inhabitants of South Africa, keeping systematic records of the deposits in which the various implements occurred … I have little doubt that some of our French colleagues, amidst all the dangers and anxieties which attend the present war, will avail themselves of the opportunities presented by the extensive trenching operations in northern France to extend further our knowledge of prehistoric times.
From *Nature* 10 December 1914

Figure 1 | Electron density distribution in two-dimensional materials. a, The densities of the electron clouds around hexagonal boron nitride (hBN), graphene and molybdenum disulphide (MoS_2) reveal successively lower 'porosities'. These porosities correspond to the ability of the materials to conduct protons[2]: hBN conducts better than graphene, whereas MoS_2 does not conduct. Nitrogen atoms, dark blue; boron, pink; carbon, light blue; molybdenum, brown; sulphur, yellow. **b**, The lattice structure of multi-layered hBN is aligned (as are its 'pores'), whereas that of multi-layered graphene is staggered so that its pores are not above each other; double-headed arrows indicate atoms sandwiched between pores. This explains why bi- and trilayers of hBN conduct protons, but bilayers of graphene do not.

two-dimensional materials. In monolayers, the electron clouds of hBN are more 'porous' than those of graphene (Fig. 1). MoS_2 does not have any 'pores' in its electron cloud, and so does not conduct protons. In multilayered hBN, the pores of successive layers align with each other, allowing protons to pass. By contrast, the lattice in multi-layered graphene is staggered such that the electron cloud of one layer blocks the pores in the next layer.

The proton conductivity of both graphene and hBN exhibited Arrhenius-type exponential increases with temperature, but graphene showed a faster rate of increase than hBN. Such temperature-dependent behaviour indicates that proton transport involves passage across an energy barrier, rather than some other mechanism. Hu and co-workers also showed that the proton conductivity could be enhanced more than tenfold by simply coating the two-dimensional materials with a discontinuous layer of platinum, a widely used catalyst often found in fuel cells.

Proton-conductive membranes are at the heart of proton-exchange membrane fuel cells, in which the 'proton exchange' membrane must conduct protons while preventing crossover of water and methanol[3]. Considerable efforts have been directed towards developing moisture-free membranes that can operate at high temperatures (greater than 120 °C) to resolve several technical problems and improve fuel-cell performance, but no membrane has completely succeeded in replacing conventional, low-temperature hydrated membranes[3]. Could graphene or hBN — which exhibit high proton conductivity but are otherwise impenetrable — provide the long-sought solution? Graphene monolayers are stable in oxygen up to 400 °C (ref. 4), whereas hBN is even more stable (its nanotube form survives temperatures of 700 °C in air[5]). And in Hu and colleagues' experiments, platinum-coated hBN was so conductive that it was essentially 'invisible' to protons. In all likelihood, the proton conductivities of pristine graphene and platinum-coated hBN exceed 50 siemens per square centimetre at high temperatures — this is the target[6] set by the US Department of Energy for the conductance of proton-exchange membranes to be developed by the year 2020. However, it may be difficult to create the large membranes of pristine graphene or hBN needed for fuel-cell applications. One practical solution could be to make a composite membrane of graphene or hBN flakes and a platinum catalyst, along the lines of another fuel-cell membrane made from flakes of graphene oxide that was reported this year[7].

The electrical properties of graphene and hBN are diametrically opposed — which, in the context of Hu and co-workers' findings, means that graphene is an electrically conductive proton conductor, whereas hBN is an electrically insulating proton conductor. The insulating characteristics of hBN raise the intriguing possibility of creating ultrathin

fuel cells in which the two cell electrodes are directly deposited on opposite sides of hBN. By contrast, the conductive properties of graphene might allow the flow of protons through it to be modulated by a gating voltage, or enable graphene to act as both a selective membrane and an electrode. Indeed, Hu *et al.* showed that a pure stream of hydrogen can be produced by applying a voltage to graphene that has protons on one side and a vacuum on the other. If a conventional electrode had been used, the hydrogen would have been contaminated with water vapour and dissolved gases.

The authors' results pose fundamental questions regarding transport across atomically thin two-dimensional materials. The exact mechanism of proton transport across graphene and hBN is yet to be unravelled. Further work is needed to predict proton conduction quantitatively and to understand the effects of platinum, the chemical environment and gate voltage in modulating proton transport.

Other areas of research now ripe for exploration include the interplay between conduction of protons and electrons; the behaviour of graphene as a combined membrane and electrode separating two different solutions on either side (a fundamentally new membrane–electrode combination); the transport of low-energy subatomic particles, isotopes or ions[8] across two-dimensional materials; the effects of modifying conventional electrodes by coating their surfaces with two-dimensional materials; and reactions involving proton transfer across two-dimensional materials between compounds less than one nanometre apart. Such research promises fresh insight into the nature of transport across two-dimensional materials, and opens up fascinating opportunities for tailoring these materials and their van der Waals heterostructures[9] — in which isolated atomic layers are assembled layer by layer in a given sequence — to obtain interesting functionalities. ∎

Rohit N. Karnik *is in the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.*
e-mail: karnik@mit.edu

1. Bunch, J. S. *et al. Nano Lett.* **8,** 2458–2462 (2008).
2. Hu, S. *et al. Nature* **516,** 227–230 (2014).
3. Devanathan, R. *Energ. Environ. Sci.* **1,** 101–119 (2008).
4. Liu, L. *et al. Nano Lett.* **8,** 1965–1970 (2008).
5. Chen, Y., Zou, J., Campbell, S. J. & Le Caer, G. *Appl. Phys. Lett.* **84,** 2430 (2004).
6. *Multi-Year Research, Development and Demonstration Plan* http://energy.gov/sites/prod/files/2014/03/f12/fuel_cells.pdf (US Dept Energy, 2012).
7. Paneri, A. *et al. J. Membr. Sci.* **467,** 217–225 (2014).
8. Garaj, S. *et al. Nature* **467,** 190–193 (2010).
9. Geim, A. K. & Grigorieva, I. V. *Nature* **499,** 419–425 (2013).

This article was published online on 26 November 2014.

A beacon for bacterial tubulin

The protein FtsZ forms a ring structure that constricts to allow bacterial cells to divide. A second protein, MapZ, has now been found to guide FtsZ to the correct mid–cell position in the bacterium *Streptococcus pneumoniae*. SEE LETTER P.259

ELIZABETH J. HARRY

The discovery of the highly evolutionarily conserved bacterial protein FtsZ several decades ago marked the beginning of our understanding of how a bacterial cell divides. FtsZ, an evolutionary precursor of the protein tubulin in multicellular organisms, self-polymerizes to form a structure called the FtsZ ring at the site where cell fission occurs. This process is thought to mark the earliest step in cell division. But how is the FtsZ ring correctly positioned to ensure equal partitioning of the parental cell's DNA into the two daughter cells? Although various models for how this occurs have been proposed, the mechanisms are far from fully resolved. In a paper published in this issue (page 259), Fleurie *et al.*[1] reveal that, in the human pathogen *Streptococcus pneumoniae*, the FtsZ ring is positioned by the protein MapZ, which acts as a beacon to identify the site of division.

Bacterial cells divide by forming a wall-like structure called a septum, which is composed of cell wall and cell membrane; this then splits down the middle to produce two new cells. During division, the FtsZ ring recruits at least 20 other proteins to the division site, leading to subsequent FtsZ-ring constriction and division[2,3]. Until recently, cell division had been intensively studied in only a few usually non-pathogenic bacterial species, such as *Escherichia coli* and *Bacillus subtilis*. Research on these rod-shaped bacteria led to a model in which division-site placement in bacteria is regulated by a combination of two mechanisms, known as the Min and nucleoid-occlusion systems. These systems allow division to occur only at the cell centre (mid-cell) by preventing FtsZ-ring formation at all other positions[4].

However, several pathogenic and non-pathogenic bacteria do not have Min or nucleoid-occlusion systems (some have one but not the other). Furthermore, even in bacteria that have both systems, FtsZ rings can form at mid-cell with the same precision when these systems are rendered inactive[5,6]. Research published in the past few years on FtsZ-ring positioning in other bacteria has uncovered negative and positive signalling systems that act on FtsZ-ring assembly[7]. Fleurie and colleagues' identification of MapZ (mid-cell-anchored protein Z) as being involved in the positioning of the division site in *S. pneumoniae* means that it is the first protein shown to have such a function in this

Figure 1 | Division-site selection in *Streptococcus pneumoniae*. a, According to Fleurie and colleagues' model[1] of cell division in *S. pneumoniae*, rings formed of the MapZ (green) and FtsZ (red) proteins are localized at the division site (mid-cell). **b,** The MapZ ring then splits in two, and these rings migrate from mid-cell to the future division sites of daughter cells as a result of elongation-specific cell-wall synthesis (blue), whereas the FtsZ ring remains at mid-cell. **c,** A third MapZ ring appears at mid-cell and the FtsZ ring also splits such that two additional FtsZ rings migrate to the two outer MapZ rings. **d,** Both mid-cell rings (FtsZ and MapZ) constrict to allow the cells to divide. This division is accompanied by division-specific cell-wall synthesis (purple). **e,** This results in two daughter cells, each with MapZ and FtsZ rings located to their division sites.
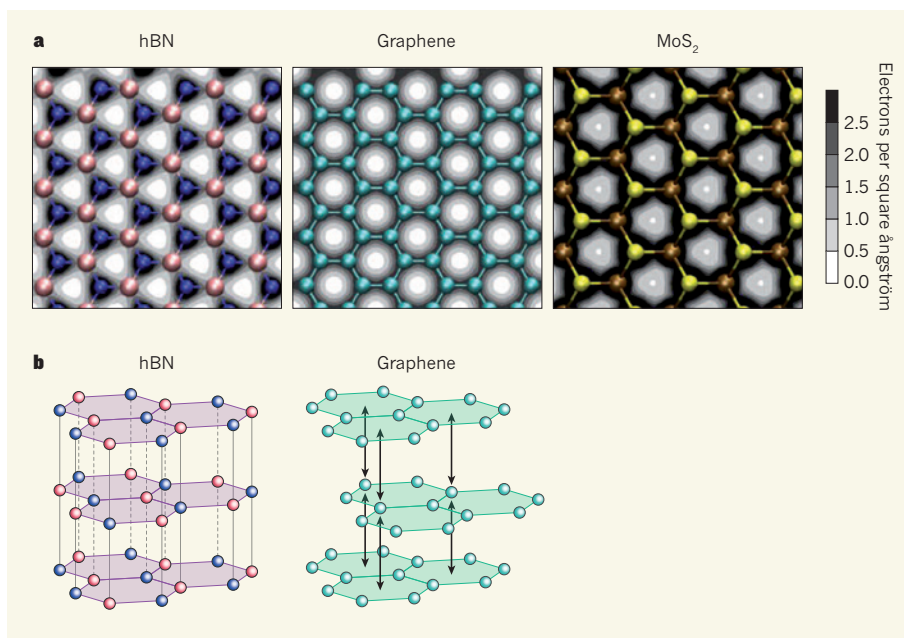
Figure 1 | Electron density distribution in two-dimensional materials. **a**, The densities of the electron clouds around hexagonal boron nitride (hBN), graphene and molybdenum disulphide (MoS_2) reveal successively lower 'porosities'. These porosities correspond to the ability of the materials to conduct protons[2]: hBN conducts better than graphene, whereas MoS_2 does not conduct. Nitrogen atoms, dark blue; boron, pink; carbon, light blue; molybdenum, brown; sulphur, yellow. **b**, The lattice structure of multi-layered hBN is aligned (as are its 'pores'), whereas that of multi-layered graphene is staggered so that its pores are not above each other; double-headed arrows indicate atoms sandwiched between pores. This explains why bi- and trilayers of hBN conduct protons, but bilayers of graphene do not.

two-dimensional materials. In monolayers, the electron clouds of hBN are more 'porous' than those of graphene (Fig. 1). MoS_2 does not have any 'pores' in its electron cloud, and so does not conduct protons. In multilayered hBN, the pores of successive layers align with each other, allowing protons to pass. By contrast, the lattice in multi-layered graphene is staggered such that the electron cloud of one layer blocks the pores in the next layer.

The proton conductivity of both graphene and hBN exhibited Arrhenius-type exponential increases with temperature, but graphene showed a faster rate of increase than hBN. Such temperature-dependent behaviour indicates that proton transport involves passage across an energy barrier, rather than some other mechanism. Hu and co-workers also showed that the proton conductivity could be enhanced more than tenfold by simply coating the two-dimensional materials with a discontinuous layer of platinum, a widely used catalyst often found in fuel cells.

Proton-conductive membranes are at the heart of proton-exchange membrane fuel cells, in which the 'proton exchange' membrane must conduct protons while preventing crossover of water and methanol[3]. Considerable efforts have been directed towards developing moisture-free membranes that can operate at high temperatures (greater than 120 °C) to resolve several technical problems and improve fuel-cell performance, but no membrane has completely succeeded in replacing conventional, low-temperature hydrated membranes[3]. Could graphene or hBN — which exhibit high proton conductivity but are otherwise impenetrable — provide the long-sought solution? Graphene monolayers are stable in oxygen up to 400 °C (ref. 4), whereas hBN is even more stable (its nanotube form survives temperatures of 700 °C in air[5]). And in Hu and colleagues' experiments, platinum-coated hBN was so conductive that it was essentially 'invisible' to protons. In all likelihood, the proton conductivities of pristine graphene and platinum-coated hBN exceed 50 siemens per square centimetre at high temperatures — this is the target[6] set by the US Department of Energy for the conductance of proton-exchange membranes to be developed by the year 2020. However, it may be difficult to create the large membranes of pristine graphene or hBN needed for fuel-cell applications. One practical solution could be to make a composite membrane of graphene or hBN flakes and a platinum catalyst, along the lines of another fuel-cell membrane made from flakes of graphene oxide that was reported this year[7].

The electrical properties of graphene and hBN are diametrically opposed — which, in the context of Hu and co-workers' findings, means that graphene is an electrically conductive proton conductor, whereas hBN is an electrically insulating proton conductor. The insulating characteristics of hBN raise the intriguing possibility of creating ultrathin

fuel cells in which the two cell electrodes are directly deposited on opposite sides of hBN. By contrast, the conductive properties of graphene might allow the flow of protons through it to be modulated by a gating voltage, or enable graphene to act as both a selective membrane and an electrode. Indeed, Hu *et al.* showed that a pure stream of hydrogen can be produced by applying a voltage to graphene that has protons on one side and a vacuum on the other. If a conventional electrode had been used, the hydrogen would have been contaminated with water vapour and dissolved gases.

The authors' results pose fundamental questions regarding transport across atomically thin two-dimensional materials. The exact mechanism of proton transport across graphene and hBN is yet to be unravelled. Further work is needed to predict proton conduction quantitatively and to understand the effects of platinum, the chemical environment and gate voltage in modulating proton transport.

Other areas of research now ripe for exploration include the interplay between conduction of protons and electrons; the behaviour of graphene as a combined membrane and electrode separating two different solutions on either side (a fundamentally new membrane–electrode combination); the transport of low-energy subatomic particles, isotopes or ions[8] across two-dimensional materials; the effects of modifying conventional electrodes by coating their surfaces with two-dimensional materials; and reactions involving proton transfer across two-dimensional materials between compounds less than one nanometre apart. Such research promises fresh insight into the nature of transport across two-dimensional materials, and opens up fascinating opportunities for tailoring these materials and their van der Waals heterostructures[9] — in which isolated atomic layers are assembled layer by layer in a given sequence — to obtain interesting functionalities. ∎

Rohit N. Karnik *is in the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.*
e-mail: karnik@mit.edu

1. Bunch, J. S. *et al. Nano Lett.* **8,** 2458–2462 (2008).
2. Hu, S. *et al. Nature* **516,** 227–230 (2014).
3. Devanathan, R. *Energ. Environ. Sci.* **1,** 101–119 (2008).
4. Liu, L. *et al. Nano Lett.* **8,** 1965–1970 (2008).
5. Chen, Y., Zou, J., Campbell, S. J. & Le Caer, G. *Appl. Phys. Lett.* **84,** 2430 (2004).
6. *Multi-Year Research, Development and Demonstration Plan* http://energy.gov/sites/prod/files/2014/03/f12/fuel_cells.pdf (US Dept Energy, 2012).
7. Paneri, A. *et al. J. Membr. Sci.* **467,** 217–225 (2014).
8. Garaj, S. *et al. Nature* **467,** 190–193 (2010).
9. Geim, A. K. & Grigorieva, I. V. *Nature* **499,** 419–425 (2013).

A beacon for bacterial tubulin

The protein FtsZ forms a ring structure that constricts to allow bacterial cells to divide. A second protein, MapZ, has now been found to guide FtsZ to the correct mid-cell position in the bacterium *Streptococcus pneumoniae*. SEE LETTER P.259

ELIZABETH J. HARRY

The discovery of the highly evolutionarily conserved bacterial protein FtsZ several decades ago marked the beginning of our understanding of how a bacterial cell divides. FtsZ, an evolutionary precursor of the protein tubulin in multicellular organisms, self-polymerizes to form a structure called the FtsZ ring at the site where cell fission occurs. This process is thought to mark the earliest step in cell division. But how is the FtsZ ring correctly positioned to ensure equal partitioning of the parental cell's DNA into the two daughter cells? Although various models for how this occurs have been proposed, the mechanisms are far from fully resolved. In a paper published in this issue (page 259), Fleurie *et al.*[1] reveal that, in the human pathogen *Streptococcus pneumoniae*, the FtsZ ring is positioned by the protein MapZ, which acts as a beacon to identify the site of division.

Bacterial cells divide by forming a wall-like structure called a septum, which is composed of cell wall and cell membrane; this then splits down the middle to produce two new cells. During division, the FtsZ ring recruits at least 20 other proteins to the division site, leading to subsequent FtsZ-ring constriction and division[2,3]. Until recently, cell division had been intensively studied in only a few usually non-pathogenic bacterial species, such as *Escherichia coli* and *Bacillus subtilis*. Research on these rod-shaped bacteria led to a model in which division-site placement in bacteria is regulated by a combination of two mechanisms, known as the Min and nucleoid-occlusion systems. These systems allow division to occur only at the cell centre (mid-cell) by preventing FtsZ-ring formation at all other positions[4].

However, several pathogenic and non-pathogenic bacteria do not have Min or nucleoid-occlusion systems (some have one but not the other). Furthermore, even in bacteria that have both systems, FtsZ rings can form at mid-cell with the same precision when these systems are rendered inactive[5,6]. Research published in the past few years on FtsZ-ring positioning in other bacteria has uncovered negative and positive signalling systems that act on FtsZ-ring assembly[7]. Fleurie and colleagues' identification of MapZ (mid-cell-anchored protein Z) as being involved in the positioning of the division site in *S. pneumoniae* means that it is the first protein shown to have such a function in this

Figure 1 | Division-site selection in *Streptococcus pneumoniae*. a, According to Fleurie and colleagues' model[1] of cell division in *S. pneumoniae*, rings formed of the MapZ (green) and FtsZ (red) proteins are localized at the division site (mid-cell). **b**, The MapZ ring then splits in two, and these rings migrate from mid-cell to the future division sites of daughter cells as a result of elongation-specific cell-wall synthesis (blue), whereas the FtsZ ring remains at mid-cell. **c**, A third MapZ ring appears at mid-cell and the FtsZ ring also splits such that two additional FtsZ rings migrate to the two outer MapZ rings. **d**, Both mid-cell rings (FtsZ and MapZ) constrict to allow the cells to divide. This division is accompanied by division-specific cell-wall synthesis (purple). **e**, This results in two daughter cells, each with MapZ and FtsZ rings located to their division sites.

oval-shaped bacterium, which has neither Min nor nucleoid-occlusion systems.

The authors demonstrate that deletion of the *mapZ* gene leads to misplacement of the FtsZ ring and the division septum, which are normally positioned at mid-cell. Using time-lapse and three-dimensional structured illumination microscopy, they show that MapZ precedes the FtsZ ring in localizing to mid-cell, and that MapZ also forms a ring structure (Fig. 1). Particularly intriguing is their finding that, once both the MapZ ring and FtsZ ring locate to mid-cell, the MapZ ring splits into two and moves to the two future division sites, whereas the FtsZ ring stays at mid-cell. A third MapZ ring then forms at mid-cell. The FtsZ ring subsequently splits and two rings migrate to co-localize with the two outer MapZ rings. The mid-cell MapZ/FtsZ rings then close to complete cell division.

How do the MapZ rings migrate to the future division sites? The occurrence of MapZ is restricted to the Streptococcaceae family and most other families of the order Lactobacillales. Fleurie and colleagues show that the MapZ-ring migration relates to the distinct mode of cell elongation in these oval-shaped cells. To increase cell size, cell-wall synthesis in streptococci begins at the mid-cell division site and moves in both directions towards the future division sites[8]. This is in contrast to elongation in rod-shaped cells, which involves cell-wall synthesis along all of the long axis of the cell. By studying fluorescently labelled cell-wall substrates and MapZ in live *S. pneumoniae* cells, the authors show that MapZ-ring migration is coupled to cell-wall synthesis. This result was further supported by their finding that MapZ binds to the cell-wall material peptidoglycan, and that inhibition of cell-wall synthesis using the antibiotic vancomycin delocalizes MapZ.

What remained to be shown was direct evidence that MapZ functions to guide FtsZ to the division site. The authors provide this by demonstrating a direct interaction between FtsZ and MapZ, which seems to depend on the 41 amino-acid residues at the amino terminus of MapZ. Further, they showed that deletion of this region still allowed MapZ septal localization but caused FtsZ to be delocalized. These studies were complicated by the fact that *mapZ*-deficient cells are often misshapen, but the authors' inspection of FtsZ-ring positioning in normal-shaped *mapZ*-deficient cells supported these results.

MapZ was first identified as Spr0334, a protein with no assigned function but that was known to be phosphorylated by *S. pneumoniae* StkP, a kinase protein involved in septum assembly, maintaining cell shape and localization of cell-wall synthesis[9]. Fleurie *et al.* show that phosphorylation of MapZ occurs at two threonine amino-acid residues at positions 67 and 78, not in the region of MapZ that directly interacts with FtsZ. They also show that non-phosphorylated MapZ still interacts with FtsZ,

and that the lack of phosphorylation does not affect Z-ring positioning. It seems, therefore, that the phosphorylation state of MapZ is important for another regulatory role for the protein, possibly in the splitting, stability and constriction of the FtsZ ring.

Putting this information together, the authors predict that MapZ has a single transmembrane anchor region that links a cytoplasmic amino-terminal domain to an extracellular carboxy-terminal domain. The extracellular domain binds peptidoglycan, thereby localizing MapZ to the division site, and the cytoplasmic domain acts as the beacon for FtsZ. As cell-wall synthesis occurs, MapZ remains attached, to arrive at the new cell equators of the forming daughter cells. The authors also propose that the phosphorylation of MapZ by StkP (and its dephosphorylation by the enzyme PhpP) regulates FtsZ-ring constriction by direct interaction with other division proteins at mid-cell, not with FtsZ. Although details of these processes remain to be determined, the findings already show that division-site positioning in bacteria is surprisingly diverse — perhaps a consequence of the diversity of lifestyle, cell shape and mode of cell-wall synthesis in these organisms.

Fleurie *et al.* also found that FtsZ forms aberrant, non-ring structures in *mapZ*-deleted cells. The authors suggest that the abnormal cell-wall synthesis and morphology of these mutant

cells is a consequence of these abnormal FtsZ-ring structures. But perhaps it is the other way around, and this idea would be worth testing, particularly in light of accumulating support for peptidoglycan structures ('piecrusts') that are proposed to mark the future FtsZ-ring assembly site in some bacteria[10]. Exploring this idea may answer the question of how MapZ itself is localized to the division site, and thus just what is the first step in bacterial cell division. ∎

Elizabeth J. Harry *is at the ithree institute, University of Technology Sydney, Sydney, New South Wales 2007, Australia.*
e-mail: liz.harry@uts.edu.au

1. Fleurie, A. *et al. Nature* **516,** 259–262 (2014).
2. Adams, D. W. & Errington, J. *Nature Rev. Microbiol.* **7,** 642–653 (2009).
3. de Boer, P. A. J. *Curr. Opin. Microbiol.* **13,** 730–737 (2010).
4. Lutkenhaus, J. *Annu. Rev. Biochem.* **76,** 539–562 (2007).
5. Rodrigues, C. D. A. & Harry, E. J. *PLoS Genet.* **8,** e1002561 (2012).
6. Bailey, M. W., Bisicchia, P., Warren. B. T., Sherratt, D. J. & Männik, J. *PLoS Genet.* **10,** e1004504 (2014).
7. Monahan, L. G., Liew, A. T. F., Bottomley, A. L. & Harry, E. J. *Front. Microbiol.* **5,** 19 (2014).
8. Pinho, M. G., Kjos, M. & Veening, J.-W. *Nature Rev. Microbiol.* **11,** 601–614 (2013).
9. Beilharz, K. *et al. Proc. Natl Acad. Sci. USA* **109,** E905–E913 (2012).
10. Turner, R. D., Vollmer, W. & Foster, S. J. *Mol. Microbiol.* **91,** 862–874 (2014).

This article was published online on 26 November 2014.

Calcium–activated proteins visualized

The first crystal structures of bestrophin and lipid scramblase proteins cast light on how these protein families transport very different substrates across membranes, yet are both activated by calcium ions. SEE ARTICLES P.207 & P.213

MATT WHORTON

Blood clotting, olfaction and vision are just a few of the many physiological processes regulated by proteins called calcium-activated chloride channels and lipid scramblases. These proteins reside in the cell membrane and control the transport of chloride ions and lipids across this otherwise impermeable barrier. The machinery underlying these activities remained unknown for decades until research by several groups over the past 12 years showed that they are at least partially comprised of the TMEM16 and bestrophin families of proteins[1–7]. In two papers published in this issue, Brunner *et al.*[8] (page 207) and Kane Dickson *et al.*[9] (page 213) report the breakthrough determination of the

three-dimensional structures of two of these proteins.

Calcium is a ubiquitous signalling ion. Certain events, such as the activation of cell-surface receptors by hormones or neurotransmitter molecules, can lead to a rapid increase in the intracellular calcium concentration. The extra calcium ions can then interact with and regulate different types of protein, including ion channels and transporters.

When calcium-activated chloride channels (CaCCs) bind calcium, they open to let chloride ions flow through the cell membrane. Moving ions across the membrane changes the electrical properties of a cell, and this can in turn affect other cellular functions. For example, CaCCs have been shown to affect the firing rate of neurons and to modulate the sensation

oval-shaped bacterium, which has neither Min nor nucleoid-occlusion systems.

The authors demonstrate that deletion of the *mapZ* gene leads to misplacement of the FtsZ ring and the division septum, which are normally positioned at mid-cell. Using time-lapse and three-dimensional structured illumination microscopy, they show that MapZ precedes the FtsZ ring in localizing to mid-cell, and that MapZ also forms a ring structure (Fig. 1). Particularly intriguing is their finding that, once both the MapZ ring and FtsZ ring locate to mid-cell, the MapZ ring splits into two and moves to the two future division sites, whereas the FtsZ ring stays at mid-cell. A third MapZ ring then forms at mid-cell. The FtsZ ring subsequently splits and two rings migrate to co-localize with the two outer MapZ rings. The mid-cell MapZ/FtsZ rings then close to complete cell division.

How do the MapZ rings migrate to the future division sites? The occurrence of MapZ is restricted to the Streptococcaceae family and most other families of the order Lactobacillales. Fleurie and colleagues show that the MapZ-ring migration relates to the distinct mode of cell elongation in these oval-shaped cells. To increase cell size, cell-wall synthesis in streptococci begins at the mid-cell division site and moves in both directions towards the future division sites[8]. This is in contrast to elongation in rod-shaped cells, which involves cell-wall synthesis along all of the long axis of the cell. By studying fluorescently labelled cell-wall substrates and MapZ in live *S. pneumoniae* cells, the authors show that MapZ-ring migration is coupled to cell-wall synthesis. This result was further supported by their finding that MapZ binds to the cell-wall material peptidoglycan, and that inhibition of cell-wall synthesis using the antibiotic vancomycin delocalizes MapZ.

What remained to be shown was direct evidence that MapZ functions to guide FtsZ to the division site. The authors provide this by demonstrating a direct interaction between FtsZ and MapZ, which seems to depend on the 41 amino-acid residues at the amino terminus of MapZ. Further, they showed that deletion of this region still allowed MapZ septal localization but caused FtsZ to be delocalized. These studies were complicated by the fact that *mapZ*-deficient cells are often misshapen, but the authors' inspection of FtsZ-ring positioning in normal-shaped *mapZ*-deficient cells supported these results.

MapZ was first identified as Spr0334, a protein with no assigned function but that was known to be phosphorylated by *S. pneumoniae* StkP, a kinase protein involved in septum assembly, maintaining cell shape and localization of cell-wall synthesis[9]. Fleurie *et al.* show that phosphorylation of MapZ occurs at two threonine amino-acid residues at positions 67 and 78, not in the region of MapZ that directly interacts with FtsZ. They also show that non-phosphorylated MapZ still interacts with FtsZ,

and that the lack of phosphorylation does not affect Z-ring positioning. It seems, therefore, that the phosphorylation state of MapZ is important for another regulatory role for the protein, possibly in the splitting, stability and constriction of the FtsZ ring.

Putting this information together, the authors predict that MapZ has a single transmembrane anchor region that links a cytoplasmic amino-terminal domain to an extracellular carboxy-terminal domain. The extracellular domain binds peptidoglycan, thereby localizing MapZ to the division site, and the cytoplasmic domain acts as the beacon for FtsZ. As cell-wall synthesis occurs, MapZ remains attached, to arrive at the new cell equators of the forming daughter cells. The authors also propose that the phosphorylation of MapZ by StkP (and its dephosphorylation by the enzyme PhpP) regulates FtsZ-ring constriction by direct interaction with other division proteins at mid-cell, not with FtsZ. Although details of these processes remain to be determined, the findings already show that division-site positioning in bacteria is surprisingly diverse — perhaps a consequence of the diversity of lifestyle, cell shape and mode of cell-wall synthesis in these organisms.

Fleurie *et al.* also found that FtsZ forms aberrant, non-ring structures in *mapZ*-deleted cells. The authors suggest that the abnormal cell-wall synthesis and morphology of these mutant

cells is a consequence of these abnormal FtsZ-ring structures. But perhaps it is the other way around, and this idea would be worth testing, particularly in light of accumulating support for peptidoglycan structures ('piecrusts') that are proposed to mark the future FtsZ-ring assembly site in some bacteria[10]. Exploring this idea may answer the question of how MapZ itself is localized to the division site, and thus just what is the first step in bacterial cell division. ∎

Elizabeth J. Harry *is at the ithree institute, University of Technology Sydney, Sydney, New South Wales 2007, Australia. e-mail: liz.harry@uts.edu.au*

1. Fleurie, A. *et al. Nature* **516,** 259–262 (2014).
2. Adams, D. W. & Errington, J. *Nature Rev. Microbiol.* **7,** 642–653 (2009).
3. de Boer, P. A. J. *Curr. Opin. Microbiol.* **13,** 730–737 (2010).
4. Lutkenhaus, J. *Annu. Rev. Biochem.* **76,** 539–562 (2007).
5. Rodrigues, C. D. A. & Harry, E. J. *PLoS Genet.* **8,** e1002561 (2012).
6. Bailey, M. W., Bisicchia, P., Warren. B. T., Sherratt, D. J. & Männik, J. *PLoS Genet.* **10,** e1004504 (2014).
7. Monahan, L. G., Liew, A. T. F., Bottomley, A. L. & Harry, E. J. *Front. Microbiol.* **5,** 19 (2014).
8. Pinho, M. G., Kjos, M. & Veening, J.-W. *Nature Rev. Microbiol.* **11,** 601–614 (2013).
9. Beilharz, K. *et al. Proc. Natl Acad. Sci. USA* **109,** E905–E913 (2012).
10. Turner, R. D., Vollmer, W. & Foster, S. J. *Mol. Microbiol.* **91,** 862–874 (2014).

This article was published online on 26 November 2014.

Calcium–activated proteins visualized

The first crystal structures of bestrophin and lipid scramblase proteins cast light on how these protein families transport very different substrates across membranes, yet are both activated by calcium ions. SEE ARTICLES P.207 & P.213
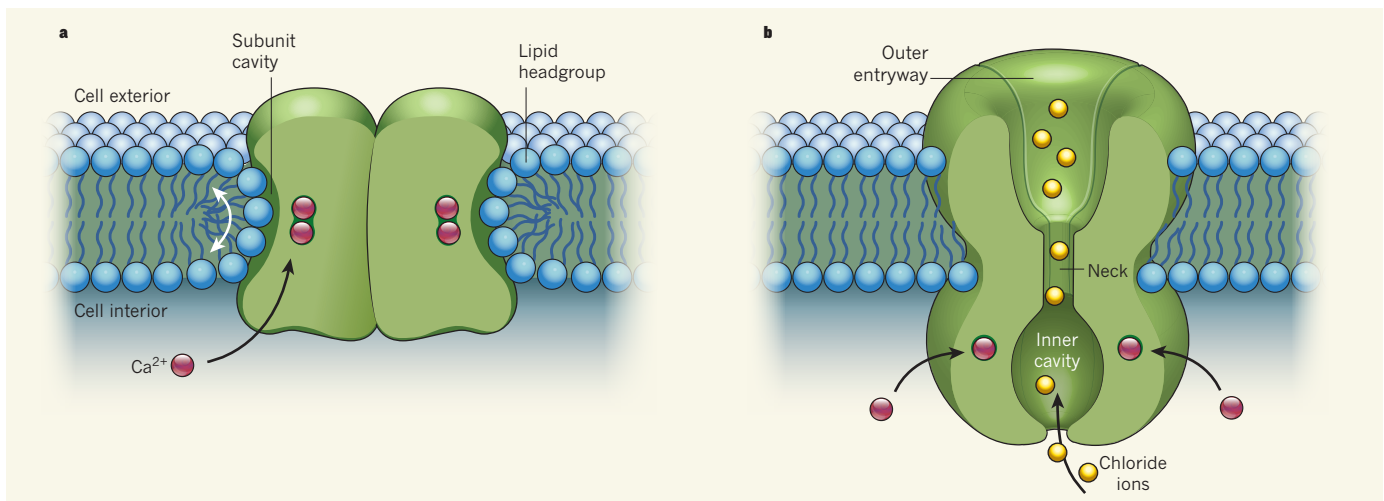
MATT WHORTON

Blood clotting, olfaction and vision are just a few of the many physiological processes regulated by proteins called calcium-activated chloride channels and lipid scramblases. These proteins reside in the cell membrane and control the transport of chloride ions and lipids across this otherwise impermeable barrier. The machinery underlying these activities remained unknown for decades until research by several groups over the past 12 years showed that they are at least partially comprised of the TMEM16 and bestrophin families of proteins[1–7]. In two papers published in this issue, Brunner *et al.*[8] (page 207) and Kane Dickson *et al.*[9] (page 213) report the breakthrough determination of the

three-dimensional structures of two of these proteins.

Calcium is a ubiquitous signalling ion. Certain events, such as the activation of cell-surface receptors by hormones or neurotransmitter molecules, can lead to a rapid increase in the intracellular calcium concentration. The extra calcium ions can then interact with and regulate different types of protein, including ion channels and transporters.

When calcium-activated chloride channels (CaCCs) bind calcium, they open to let chloride ions flow through the cell membrane. Moving ions across the membrane changes the electrical properties of a cell, and this can in turn affect other cellular functions. For example, CaCCs have been shown to affect the firing rate of neurons and to modulate the sensation

Figure 1 | Cartoons of an activated lipid scramblase and a chloride channel. a, Brunner *et al.*[8] report the structure of nhTMEM16, a calcium-activated lipid scramblase protein found in cell membranes. They find that nhTMEM16 is a dimer, in which each identical subunit has two binding sites for calcium ions and a narrow crevice (the subunit cavity) that spans the membrane. Calcium binding opens the crevice, allowing lipid headgroups to move from one side of the membrane to the other (white arrow). **b**, Kane Dickson *et al.*[9] have solved the crystal structure of bestrophin 1 (BEST1), a chloride channel. They find that BEST1 forms a pore from five identical subunits (two are omitted here, to reveal the pore). Binding of calcium ions opens the pore, which consists of an outer entryway, a narrow neck and an inner cavity. The pore is negatively charged at the top but positively charged at the bottom, to aid anion selectivity.

of odorants in olfactory cells[10].

Bestrophins and some of the TMEM16 proteins have been classified[1–4] as CaCCs, but other members of the TMEM16 family are not ion channels. Instead, they are lipid scramblases[5,6] — proteins that let phospholipids flip from one side of the cell membrane to the other. This is crucial in many physiological processes, with one of the most established roles being that of exposing a lipid called phosphatidylserine to the surface of platelets to initiate blood clotting.

The structures reported by Brunner *et al.* and Kane Dickson *et al.* start to address the basic mechanisms of how bestrophins and TMEM16 proteins work. Namely, how does calcium binding lead to opening of a chloride-selective channel or activation of a lipid scramblase? And, in the case of the TMEM16 family, how can proteins with similar sequences (and thus structures) transport such different substrates?

Brunner *et al.* solved the structure of nhTMEM16, a calcium-activated lipid scramblase from the fungus *Nectria haematococca*. Around 40% of the amino-acid sequence of nhTMEM16 is identical to those of its mammalian counterparts, so it probably shares the same basic structure and mechanism of action. They find that the protein is organized as a dimer of two identical subunits (Fig. 1a). Each subunit has a region the authors call the subunit cavity, a narrow crevice that spans the membrane.

The cavity is lined by hydrophilic amino-acid residues, which is remarkable because it is exposed to the hydrophobic lipid environment of the cell membrane. The authors propose that this design facilitates lipid scrambling by providing a conduit for the hydrophilic lipid headgroups, while letting the hydrophobic

lipid tails remain in the membrane. Several of the amino acids that line the subunit cavity have been implicated in ion conduction and selectivity in TMEM16 chloride channels[5,11], and so the authors propose that this cavity is also the pathway for chloride ions in those proteins.

Within each subunit of nhTMEM16, at a point that corresponds to roughly one-third of the way into the cell membrane from the cytoplasm, just behind the subunit cavity, lies a binding site for two calcium ions. The location of the binding site within the membrane explains the voltage-dependence of TMEM16 proteins' calcium activation: more-positive voltages across the membrane make it easier for calcium ions to partially traverse the transmembrane electric field.

Many of the amino acids that bind the calcium ions are evolutionarily conserved throughout the TMEM16 family. When the researchers mutated these amino acids in mTMEM16A, a mouse chloride channel, they observed loss of activity. This suggests that the calcium-binding site probably regulates the activity of all TMEM16 family members.

Kane Dickson *et al.*[9] solved the structure of chicken bestrophin 1 (BEST1), which is 74% identical to human BEST1. The structure reveals a completely different architecture from that of nhTMEM16. Instead of a dimer of subunits that seem to function independently, the BEST1 channel is a pentamer in which the assembled subunits create a pore for chloride ions to pass through the middle of the protein complex (Fig. 1b). This pentameric assembly seems to be a versatile platform for proteins, because a similar architecture was observed in the recently solved structure[12] of KpBest — a bacterial (*Klebsiella pneumoniae*) protein that is distantly related to bestrophins in eukaryotes (organisms that include plants, animals

and fungi), but which is a cation channel and is not activated by calcium ions.

The authors propose several ways in which structures along the pore allow BEST1 channels to conduct only ions that have a single negative charge, such as chloride ions. First, a region called the outer entryway on the extracellular side of the protein is overall negatively charged. This will repel most anions, especially doubly charged ones. However, ten weakly positively charged pockets within this region are sufficient to draw in singly charged anions.

Further along the pore is a narrow region called the neck, which is mostly lined by hydrophobic amino-acid residues. This would exclude both anions and cations, were it not for a ring of phenylalanine residues at the narrowest part. The phenylalanines are arranged such that the small positive charge localized on the edge of their benzene rings points to the middle of the pore. This facilitates the passage of small anions, but blocks cation movement. Finally, the inner cavity of the pore, which resides on the cytoplasmic side of the protein, is highly positively charged, to help to attract anions from inside the cell. A narrow aperture just below this cavity may prevent the entry of larger anions such as proteins or nucleic acids, which would block the pore.

Kane Dickson and colleagues report that each BEST1 subunit has a calcium-binding site, termed the Ca^{2+} clasp. The site is located within the intracellular part of the channel, close to the neck region. Because of this proximity, the authors suggest that the neck might be closed when calcium ions are not bound at the site, but that calcium binding induces conformational changes in the protein that leads to opening of the neck.

For both BEST1 and nhTMEM16, structures of the calcium-free states will be necessary to

understand how calcium binding is transduced into mechanical work (to open the channel in BEST1 or activate the lipid transporter in nhT-MEM16). In both proteins, the calcium-binding site is completely buried by protein, suggesting that the calcium-free state must have an appreciably different conformation to allow calcium ions to enter the site from inside the cell. It will also be important to obtain a structure of a TMEM16 chloride channel to understand the structural basis for the subunit cavity's dichotomous nature — its ability to serve as either a lipid- or a chloride-ion conduit. ■

Matt Whorton is at the Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239-3098, USA.
e-mail: whorton@ohsu.edu

1. Yang, Y. D. et al. Nature **455**, 1210–1215 (2008).
2. Caputo, A. et al. Science **322**, 590–594 (2008).
3. Schroeder, B. C., Cheng, T., Jan, Y. N. & Jan, L. Y. Cell **134**, 1019–1029 (2008).
4. Sun, H., Tsunenari, T., Yau, K.-W. & Nathans, J. Proc. Natl Acad. Sci. USA **99**, 4008–4013 (2002).
5. Yang, H. et al. Cell **151**, 111–122 (2012).
6. Suzuki, J., Umeda, M., Sims, P. J. & Nagata, S. Nature **468**, 834–838 (2010).
7. Malvezzi, M. et al. Nature Commun. **4**, 2367; http://dx.doi.org/10.1038/ncomms3367 (2013).
8. Brunner, J. D., Lim, N. K., Schenck, S., Dürst, A. & Dutzler, R. Nature **516**, 207–212 (2014).
9. Kane Dickson, V., Pedi, L. & Long, S. B. Nature **516**, 213–218 (2014).
10. Hartzell, C., Putzier, I. & Arreola, J. Annu. Rev. Physiol. **67**, 719–758 (2005).
11. Yu, K., Duran, C., Qu, Z., Cui, Y. Y. & Hartzell, H. C. Circ. Res. **110**, 990–999 (2012).
12. Yang, T. et al. Science **346**, 355–359 (2014).

This article was published online on 12 November 2014.

APPLIED PHYSICS

The virtues of tiling

A cracked metal film on an elastic substrate has been shown to provide ultrahigh sensitivity in detecting mechanical vibrations. The result draws inspiration from principles of tiling that apply to many biological systems. See Letter p.222

PETER FRATZL

Sensing vibrations such as sound or other small movements is a fundamental requirement for many technical applications. In the natural world, communication is often based on emitting and sensing vibrations. Sound is just one example. The wandering spider *Cupiennius salei* (Fig. 1) scratches plant leaves with its mouth and abdomen. A prospective mate can sense and distinguish the resulting tiny plant vibrations using one of the world's most sensitive vibration sensors — the lyriform organ located in the spider's legs[1]. The lyriform sensor is based on a parallel arrangement of slits of different lengths, reminiscent of the arrangement of strings in a lyre[2]. On page 222 of this issue, Kang et al.[3] describe how the lyriform organ served as the inspiration for developing a vibration sensor of ultrahigh sensitivity.

Kang and colleagues' sensor is based on a 20-nanometre-thin platinum layer deposited on top of a comparatively soft polymer. The researchers introduced a series of parallel cracks into the layer, somewhat analogous to the parallel slits in the spider's organ, and applied a voltage to the device. The resulting electrical current will only flow through the metallic layer, with the cracks representing the major source of resistance to the passage

of this current. When a mechanical vibration reaches the device, the associated oscillatory motion cyclically stretches and compresses the system, resulting in the opening and closing of the cracks. On stretching, the cracks undergo geometric amplification (Fig. 2a), an effect that is related to the deformation that the cracks experience and that affects the sensitivity of the vibration sensor. The cracks effectively correspond to gaps between tiles in a tiled surface, and these gaps take up most of the deformation induced by the stretching. Hence, the breadth of the cracks increases by a much larger factor than the device as a whole. Figure 2a shows that this amplification factor is 100 if the spacing between tiles is 1% of the device's total length. If the spacing is 0.1% of the total length, the amplification is 1,000. Kang and co-workers show that, if the device is stretched by 0.5% of its total length, the change in its electrical resistivity is 450 times larger than that of an analogue system without cracks.

Tiling hard surfaces is a common process used to avoid the destruction of surfaces during deformation. Paved roads are an obvious example, in which accommodation to thermal expansion in hot summers or to freezing of wet soils in cold winters is provided by the interstices between the road stones. Covering the road with a continuous layer of concrete would inevitably lead to the formation of cracks or bulges. Continuous layers on the road became possible only with the invention of elastic bitumen coverings. But there are also many natural examples of armours and hard coatings that are composed of tiles and thus avoid the fracture that would follow even a small deformation. This is true for the skeleton of sharks, which consists of (relatively soft) cartilage covered with a hard layer of mineralized cartilage. Tiling of this hard layer, in the form of mineralized tesserae connected by organic fibres, avoids cracking and provides exceptional mechanical properties[4]. Tiling is also found between the ribs in the carapace of a turtle, in armoured fish scales and in many other biological materials[5,6]. These all have in common the property that splitting a hard layer into individual tiles allows for deformation by a sort of 'breathing' of the interstices between tiles, leaving

Figure 1 | The wandering spider *Cupiennius salei*. Kang et al.[3] have developed a vibration sensor whose working principle is inspired by the geometry of the lyriform sensor located in the legs of *Cupiennius salei*.
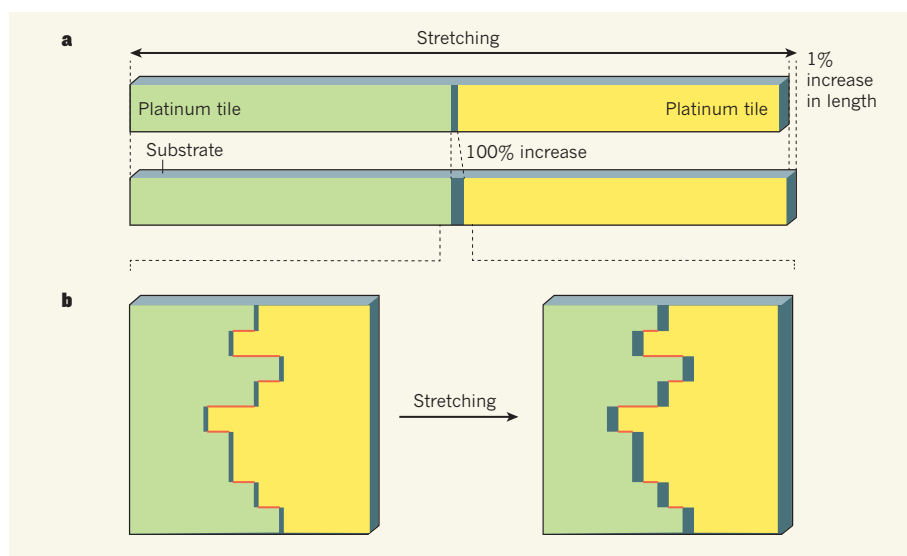
Figure 2 | Stretching rough cracks. Kang and colleagues' vibration sensor[3] is based on two effects on cracks in a platinum layer laid on a soft substrate: geometric amplification (**a**) and corrugation (**b**). When the device is stretched, the width of the crack that separates two platinum tiles increases by a much larger factor than the device as a whole, because the tiles are rigid. Shown here is a crack-width increase of 100% for a system stretched by 1%; this corresponds to an amplification factor of 100. Corrugation of the cracks at the nanoscale means that, on stretching, lateral contacts (red) between the tiles remain and allow electrical current to flow in the crack if a voltage is applied to the system. The electrical conductivity of the device is proportional to the total contact area between the tiles and thus depends on the amount of deformation.

platinum layer: their width in the nanometre range, which leads to geometric amplification, and their roughness at the nanoscale, which provides an electrical signal that depends on the amplitude of the deformation.

Kang and colleagues demonstrate that their sensor can be incorporated into devices to record minute vibrations such as musical sounds or the flapping of a ladybird's wings. Despite these impressive practical applications, the analogy with the spider's lyriform sensor is not complete. The only feature translated into the authors' system is geometric amplification. The biological sensing mechanism is entirely different (it is based on the firing of neurons rather than the measurement of electrical resistivity) and many other aspects of the spider's organ, such as its tunable sensitivity to different vibration-frequency ranges, are not reproduced. Although it may not be necessary to have these features included in a technical system, we are still far away from an artificial sensory system with a performance similar to that of the spider organ, whose evolution going back to the origins of the Chelicerata group of arthropods has been 1,000 times longer than the existence of humans. ∎

Peter Fratzl is at the Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany.
e-mail: fratzl@mpikg.mpg.de

1. Barth, F. G. *A Spider's World: Senses and Behavior* (Springer, 2002).
2. Fratzl, P. & Barth, F. G. *Nature* **462,** 442–448 (2009).
3. Kang, D. *et al. Nature* **516,** 222–226 (2014).
4. Liu, X., Dean, M. N., Summers, A. P. & Earthman, J. C. *Mater. Sci. Eng. C* **30,** 1077–1084 (2010).
5. Dunlop, J. W. C., Weinkamer, R. & Fratzl, P. *Mater. Today* **14** (3), 70–78 (2011).
6. Li, Y., Ortiz, C. & Boyce, M. C. *J. Mech. Phys. Solids* **61,** 1144–1167 (2013).

the tiles themselves essentially undeformed.

Although tiling with extremely fine interstices (generated by controlled cracking of the platinum layer) introduces geometric amplification in Kang and colleagues' device, this feature by itself does not explain how deformation during a cycle of stretching and compression is actually transformed into an electrical signal proportional to the amount of deformation. Indeed, with the idealized system sketched in Figure 2a, electrical conductivity would immediately be lost when

conducting (stiff) tiles start to separate upon stretching — that is, as soon as even the slightest deformation occurs.

In their study, Kang *et al.* take advantage of a particular property of cracks in platinum, their roughness at the nanoscale. Corrugations associated with such roughness provide lateral contacts (Fig. 2b) that enable electrical conductivity even when the gap between the platinum tiles increases. Hence, the ultrasensitivity of the sensor to vibration is due to the combination of two properties of the cracks in the

---

When wells run dry

A global analysis reveals growing societal dependence on the use of non-renewable freshwater resources that depletes groundwater reserves and undermines human resilience to water scarcity in a warming world.

RICHARD TAYLOR

That freshwater reserves are in decline in many parts of the world is not only of great scientific interest, but of profound societal concern. Reports of groundwater depletion[1,2] and declining river and lake levels[3] provide compelling evidence of regional freshwater use exceeding its renewable supply. Quantifying freshwater supply and use around the world is, however, a substantial technical

challenge. In one of the most comprehensive analyses so far, published in *Environmental Research Letters*, Wada and Bierkens[4] estimate the supply and use of fresh water from 1960 to 2099. They use both historical records and future projections that include substantial demographic and climate-related changes expected this century. Their analyses reveal a steady rise in the non-renewable use of fresh water in many parts of the world that should be of global concern.

Irrigation currently accounts for 70% of global freshwater withdrawals[4]. The green revolutions of the past half century which dramatically increased food production, most notably in the United States and Asia, were driven primarily by the expansion of cultivated land under irrigation. Because irrigation redistributes fresh water withdrawn from aquifers, rivers and lakes to the land, it changes regional water balances by increasing consumptive use of fresh water through evapotranspiration.

Intensive irrigation can deplete freshwater sources. For rivers and lakes that are being replenished through present-day precipitation, the magnitude of their depletion is constrained by their limited total volume[5] (about 93,000 cubic kilometres worldwide) and the very visible impacts of overuse. By contrast, groundwater resources derived from precipitation over years to decades and, in some cases, millennia, enable substantial non-renewable use on account of their vast, distributed
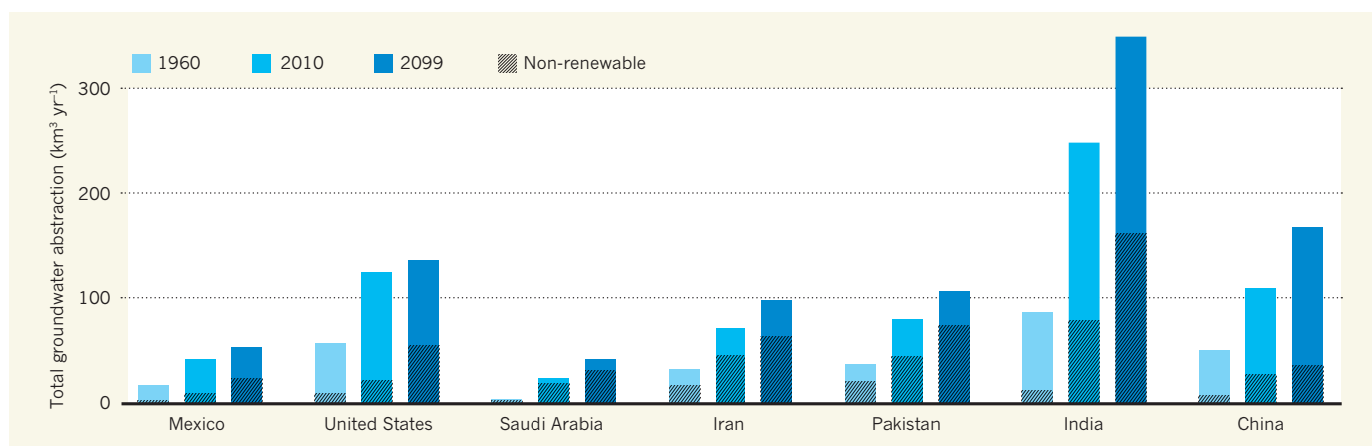
Figure 1 | Historical and projected groundwater withdrawals in the world's major irrigating countries. The chart shows total and non-renewable groundwater abstraction in India, the United States, China, Pakistan, Iran, Mexico and Saudi Arabia, as estimated by Wada and Bierkens[4], for 1960, 2010 and 2099; these countries accounted for 74% of global groundwater withdrawals in 2010. From 1960 to 2010, the estimated proportion of non-renewable groundwater withdrawals increases for all of these countries except Pakistan, where it remains stable but high at 58%.

volume[5] (about 10,500,000 km[3]) and the fact that the impacts of overuse are largely invisible. Wada and Bierkens's study marks a significant advance on previous studies because it explicitly incorporates non-renewable uses of groundwater and surface water.

From a wide range of sources, the authors compiled the most detailed estimates yet of changing agricultural, industrial and household use of fresh water from around the world. Notably, these estimates account for return flows from irrigation as well as the recycling of water from industrial and domestic withdrawals. They then compared human freshwater use to estimates of freshwater supply derived from a global hydrological model and contributions from desalinization in coastal regions. The researchers also considered future projections of freshwater supply that explicitly factor in impacts of climate change, as represented by projections from five climate models using the 'middle of the road' scenario of global warming of 4 °C by the end of this century. They then overlaid distributed freshwater supply and use to define the proportion of consumptive use that derives from non-renewable groundwater abstraction and surface-water overabstraction. Here, non-renewable groundwater abstraction is groundwater use in excess of replenishment by recharge, whereas surface-water overabstraction is defined as the quantity of environmental flows denied to aquatic ecosystems though consumptive use.

Wada and Bierkens's study reveals that non-renewable freshwater use globally rose by 50% from 1960 to 2010 primarily as a result of the expansion of irrigation in the United States, China, India, Pakistan, Mexico, Saudi Arabia and northern Iran. Crucially, this rise is primarily attributed to non-renewable groundwater withdrawals (Fig. 1). As a result, groundwater is now estimated to account for 50% of freshwater withdrawals globally. Future projections indicate that climate change will

exacerbate non-renewable freshwater use in the Mediterranean, southern Africa, the United States, Mexico and the Middle East. Globally, non-renewable freshwater use is projected to increase by one third by the end of the twenty-first century and to comprise 40% of human water consumption. This additional increase is expected to come largely from non-renewable groundwater withdrawals.

There are, however, some important limitations to this analysis. First, renewable freshwater resources in the tropics, and especially Africa, are not well represented by the global hydrological model. Simulated river discharge in some basins is two to three times greater than that observed[6] and is likely to reflect the model's systematic underestimation of tropical evapotranspiration. Second, the estimation of groundwater withdrawals does not consider how declining groundwater levels that result from the increasing non-renewability of these withdrawals raise the energy cost of bringing groundwater to the surface and allow access only to those able to afford deeper wells. Third, the production of a single future projection of freshwater supply and use based on mean output from five different climate models masks uncertainty in climate-change impacts. Fourth, the analysis does not consider water quality and how fresh water recycled from agricultural, industrial and domestic withdrawals may reduce rather than enhance freshwater supply. These limitations do not, however, undermine the robustness of the authors' central conclusion of the growing dependence of humans on the use of non-renewable freshwater resources.

Our increased use of such resources depletes groundwater storage and compromises the operation of aquatic ecosystems that sustain fisheries and other vital services. Indeed, groundwater depletion observed in some of the world's major agricultural regions[1] now threatens global food production. This depletion undermines our resilience not only to

future increases in freshwater demand[4] but also to global warming. In a warming world, precipitation is intensified, occurring in fewer but heavier rainfall events[7]. The resulting impact of longer droughts and greater variability in river discharges will amplify human reliance on stored groundwater when this resource is in decline in many regions, and on surface-water storage when most of the world's major river systems are already dammed[8].

We need to better understand available groundwater storage and recharge responses to the intensification of rainfall, which is expected to be especially strong in the tropics[7]. Indeed, it is here where increases in freshwater use are projected to be most intense[4]. We also need to reduce human dependence on non-renewable fresh water through more efficient water use, particularly in irrigation, and by trading in 'virtual water'[9], which reduces local freshwater use through the import of food and other products. If we continue along our present trajectory, "when the well runs dry we (shall) know the worth of water"[10]. ∎

Richard Taylor *is in the Department of Geography, University College London, London WC1E 6BT, UK.*
e-mail: richard.taylor@ucl.ac.uk

1. Famiglietti, J. S. *Nature Clim. Change* **4,** 945–948 (2014).
2. Gleeson, T., Wada, Y., Bierkens, M. F. P. & van Beek, L. P. H. *Nature* **488,** 197–200 (2012).
3. Pala, C. *Science* **334,** 303 (2011).
4. Wada, Y. & Bierkens, M. F. P. *Environ. Res. Lett.* **9,** 104003 (2014).
5. Shiklomanov, I. A. & Rodda, J. C. (eds) *World Water Resources at the Beginning of the 21st Century* (Cambridge Univ. Press, 2004).
6. Wada, Y., Wisser, D. & Bierkens, M. F. P. *Earth Syst. Dynam.* **5,** 15–40 (2014).
7. Allan, R. P., Soden, B. J., John, V. O., Ingram, W. & Good, P. *Environ. Res. Lett.* **5,** 025205 (2010).
8. Lehner, B. et al. *Front. Ecol. Environ.* **9,** 494–502 (2011).
9. Allan, J. A. *Groundwater* **36,** 545–546 (1998).
10. Franklin, B. *The Private Life of the Late Benjamin Franklin, LL.D* (J. Parsons, 1793).

REVIEW

Catalytic enantioselective synthesis of quaternary carbon stereocentres

Kyle W. Quasdorf[1] & Larry E. Overman[1]

Quaternary carbon stereocentres—carbon atoms to which four distinct carbon substituents are attached—are common features of molecules found in nature. However, before recent advances in chemical catalysis, there were few methods of constructing single stereoisomers of this important structural motif. Here we discuss the many catalytic enantioselective reactions developed during the past decade for the synthesis of single stereoisomers of such organic molecules. This progress now makes it possible to incorporate quaternary stereocentres selectively in many organic molecules that are useful in medicine, agriculture and potentially other areas such as flavouring, fragrances and materials.

The properties of organic molecules are intimately related to their shape. In many structurally complex organic molecules, shape is influenced—or dictated—by the three-dimensional orientation of substituents at carbon stereocentres. During the last third of the twentieth century, chemists succeeded in developing many powerful methods for directly forming a single three-dimensional orientation (configuration) of carbon centres of this type having one hydrogen substituent. In marked contrast, the construction of a single configuration of stereogenic carbon centres having four different carbon substituents (hereafter referred to as quaternary stereocentres) has until just recently been a daunting challenge for chemical synthesis. However, remarkable advances have been recorded during the past decade in the stereocontrolled construction of quaternary stereocentres using chemical catalysis.

Quaternary stereocentres are found in many biologically active small-molecule natural products, as exemplified by cortisone and morphine (Fig. 1a). One of the difficulties in constructing quaternary carbons is their congested nature, which is illustrated in the space-filling model of morphine wherein this carbon is barely visible at the end of the pointing arrow (Fig. 1b). Besides the challenge of steric hindrance, the stereoselective construction of quaternary stereocentres must involve the use of carbon–carbon bond-forming reactions that provide the desired three-dimensional orientation of the four attached substituents, that is, the correct absolute configuration of the quaternary stereocentre. The structures of current pharmaceutical agents provide one indication of the challenges involved: molecules containing a quaternary stereocentre comprised 12% of the top 200 prescription drugs sold in the US in 2011[1]. However, all of these drugs are derived from naturally occurring compounds (steroids, opioids or taxane diterpenoids), with a natural product precursor providing the quaternary stereocentres of the marketed drug in virtually every case (see, for example, ref. 2). The near absence of approved drugs containing chemically synthesized quaternary carbon stereocentres reflects the situation that until recently few reliable methods for preparing such structures existed (see ref. 3, ref. 4 (and reviews of enantioselective synthesis of quaternary stereocentres cited therein) and ref. 5).

In 2004, we surveyed the field of catalytic enantioselective synthesis of quaternary stereocentres and concluded that only four transformations—Diels–Alder reactions, reactions of chiral allylmetal intermediates with carbon nucleophiles, intramolecular Heck reactions, and reactions of chiral carbon nucleophiles with electrophiles—were well documented as useful[5]. In contrast, today a broad selection of methods is available for this purpose, prompting us to again review the status of this field. Our treatment will be

organized by general reaction type in a fashion similar to our previous review[5]. We will highlight methods for which some generality has been demonstrated, and we will concentrate on catalytic transformations whose utility has been validated by their use in the construction of complex chemical structures, typically natural products.

Cycloaddition reactions

The catalytic enantioselective construction of quaternary stereocentres by cycloaddition reactions has progressed significantly in the past decade. New catalytic paradigms have been introduced, and the type of cycloaddition
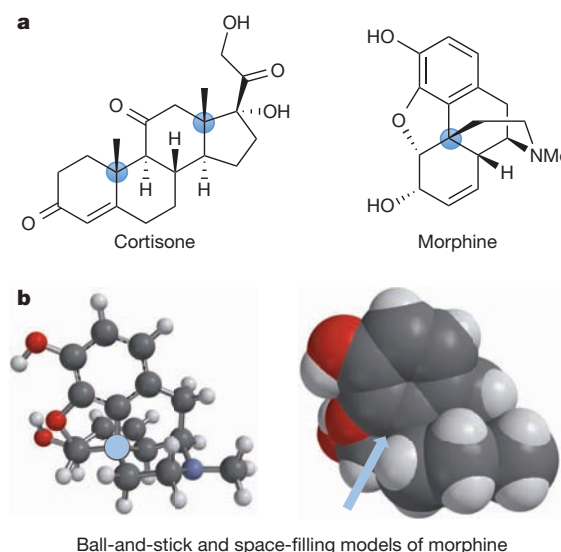
Figure 1 | **Quaternary stereocentres are important structural features of many biologically active molecules, as exemplified by the natural products cortisone and morphine. a,** Structures of the steroid cortisone and opioid morphine with their quaternary stereocentres highlighted. Me, methyl. **b,** Steric congestion, which presents a formidable challenge for chemical synthesis of molecules containing quaternary stereocentres, is illustrated in the ball-and-stick model of morphine wherein the quaternary stereocentre is highlighted by a blue circle, and particularly in the space-filling model on the right in which its sterically congested quaternary centre is barely visible at the end of the pointing arrow.

[1]Department of Chemistry, 1102 Natural Sciences II, University of California, Irvine, California 92697-2025, USA.

that can be employed has been expanded beyond classical Diels–Alder reactions.

An important recent development in this area is the use of small organic molecules to activate the dienophile in Diels–Alder reactions. The MacMillan group has described a number of [4+2]-cycloaddition reactions that proceed via catalytically generated iminium ion intermediates[6]. The utility of these reactions for the enantioselective synthesis of quaternary stereocentres was highlighted in concise total syntheses of various indole alkaloids[7,8]. For example, the rapid construction of intermediate **2** was the central step in total syntheses of (−)-minovincine (**3**), (−)-akuammicine (**4**) and (−)-strychnine (**5**) (Fig. 2a). Tetracyclic product **2** is the result of a cascade sequence, the first step of which is a catalytic enantioselective [4+2]-cycloaddition generating tricyclic intermediate **1**. Notable other recent reports of the construction of quaternary stereocentres using organocatalytic Diels–Alder reactions are the use of secondary amines[9] and hydrogen-bonding thiourea catalysts[10] to synthesize spirocyclic oxindoles and oxindole natural products.

The broad utility of catalytic enantioselective Diels–Alder reactions for constructing quaternary stereocentres is illustrated by several recent natural product total syntheses. For example, the total synthesis of *ent*-hyperforin (**10**) by the Shibasaki group featured a catalytic enantioselective Diels–Alder reaction between dienophile **6** and diene **7** in the presence of an iron complex generated from FeBr$_3$ and the pyridine bisoxazoline (PyBOX) ligand **8** (Fig. 2b)[11]. The quaternary stereocentre of cycloadduct **9** subsequently played a decisive role in evolving the two additional quaternary stereocentres of *ent*-hyperforin. Catalytic enantioselective Diels–Alder reactions that form quaternary stereocentres have also been orchestrated in various intramolecular fashions, including macrobicyclization[12] and transannular processes[13]. The former construction is illustrated in the transformation of polyene aldehyde **11** in the presence of oxazaborolidinium catalyst **12** to form macrobicyclic product **13** in good yield and 90% enantiomeric excess (e.e.). Snyder and Corey elaborated cycloadduct **13** to several natural products, including palominol (**14**) (Fig. 2c).

Catalytic enantioselective cycloaddition reactions of various types have now been used for forming quaternary stereocentres. In particular, constructions to form five-membered rings are widely established. The Davies group described the reaction of indoles with rhodium-carbenoids to produce cyclopentene-fused indolines in excellent yield and enantioselectivity (Fig. 3a)[14]. These reactions are believed to proceed in a stepwise fashion via a dipolar intermediate such as **15**. Enantioselective cycloadditions of palladium-trimethylenemethane (Pd-TMM) complexes have been developed extensively over many years by Trost and co-workers[15]. In this way, a variety of functionalized cyclopentene derivatives containing quaternary stereocentres can be accessed directly. For example, the catalytic-enantioselective cycloaddition of propylidene oxindole **16** and TMM donor **17** was the central step in the total synthesis of (−)-marcfortine C (**18**) (Fig. 3b)[16]. Several types of ligands were investigated for promoting this transformation, with phosphoramidite ligand **19** found to be optimal. Other notable examples of forming five-membered rings and quaternary stereocentres in cycloaddition reactions that employ organometallic[17] or chiral phosphoric acid catalysts have also been reported[18,19].

The formation of quaternary stereocentres by catalytic enantioselective cyclopropanation reactions was well established at the time of our previous review[5]. Progress in this area continues at a rapid pace, with the scope of enantioselective Simmons–Smith cyclopropanations and transition-metal-catalysed decomposition of diazoalkanes being continually advanced[20]. Catalytic enantioselective cyclopropanation reactions are also pivotal steps of cascade sequences developed to form larger rings. In their synthesis of (−)-5-*epi*-vibsanin E, the Davies group illustrates one variant: a cyclopropanation/Cope rearrangement sequence[21]. In the example we illustrate here, cycloaddition of the vinylcarbenoid derived from vinyl diazoester **21** and Rh$_2$(*R*-PTAD)$_4$ with the terminal double bond of diene **20** delivered *cis*-divinylcyclopropane **22**, which under the reaction conditions underwent Cope rearrangement to furnish cycloheptadiene **23** (Fig. 3c).
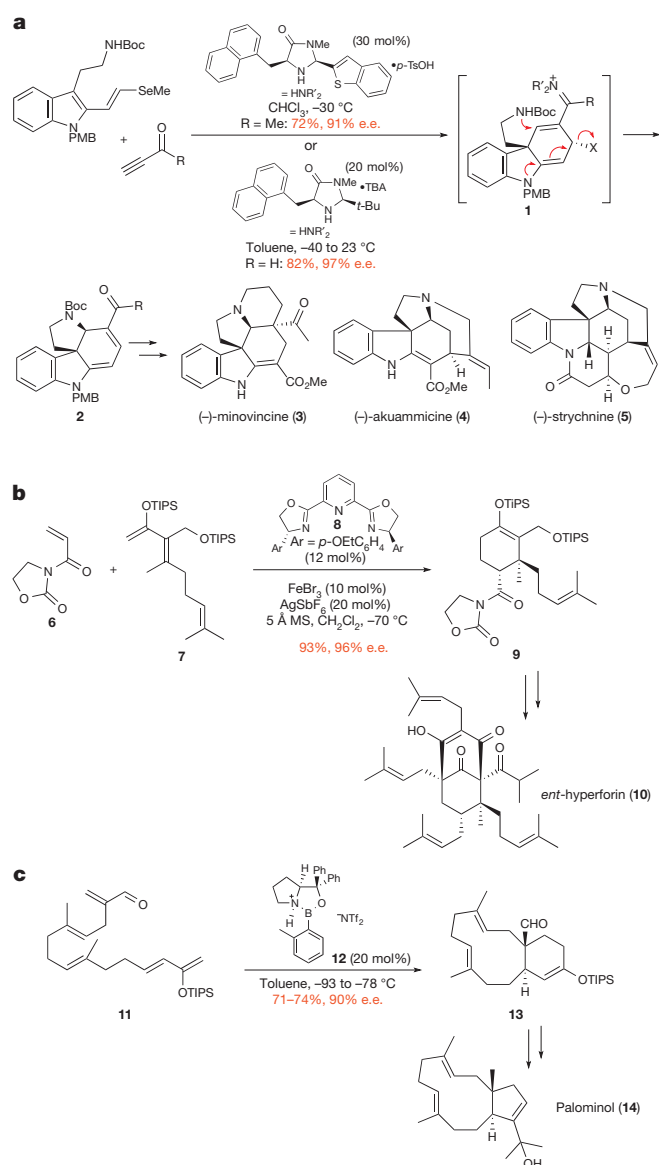
Figure 2 | The use of catalytic enantioselective Diels–Alder reactions to synthesize natural products containing quaternary stereocentres. e.e., enantiomeric excess. **a**, A bimolecular Diels–Alder reaction promoted by iminium ion activation forms intermediate **1** in the first step of a cascade sequence generating tetracyclic product **2**. This product contains the quaternary stereocentre and four rings common to several groups of indole alkaloids and was employed to complete enantioselective total syntheses of various indole alkaloids, including (−)-minovincine (**3**), (−)-akuammicine (**4**), and (−)-strychnine (**5**)[7]. Boc, *tert*-butoxycarbonyl; PMB, *p*-methoxybenzyl; *p*-TsOH, *p*-toluenesulfonic acid; *t*-Bu, *tert*-butyl; TBA, tribromoacetic acid. **b**, An iron-bisoxazoline catalysed bimolecular Diels–Alder reaction forms product **9** whose quaternary stereocentre subsequently controlled the elaboration of the two additional quaternary stereocentres of *ent*-hyperforin (**10**)[11]. TIPS, triisopropylsilyl; Et, ethyl; MS, molecular sieves. **c**, Oxazaborolidinium-catalysed intramolecular Diels–Alder reaction to form the 11-membered ring and quaternary stereocentre of palominol (**14**)[12]. Tf, trifluorosulfonyl; Ph, phenyl.

Polyene cyclizations

Enantioselective cyclization reactions of acyclic polyenes have been advanced considerably during the past decade. Yamamoto and co-workers described a number of enantioselective polyene cyclizations that proceed in the presence of stoichiometric amounts of protic acids generated upon complexation of SnCl$_4$ with BINOL-derived ligands[22] (BINOL, 1,1′-bi-2-naphthol). Building on these disclosures, Corey and co-workers reported several concise total syntheses in which polyene cyclizations promoted by complexes
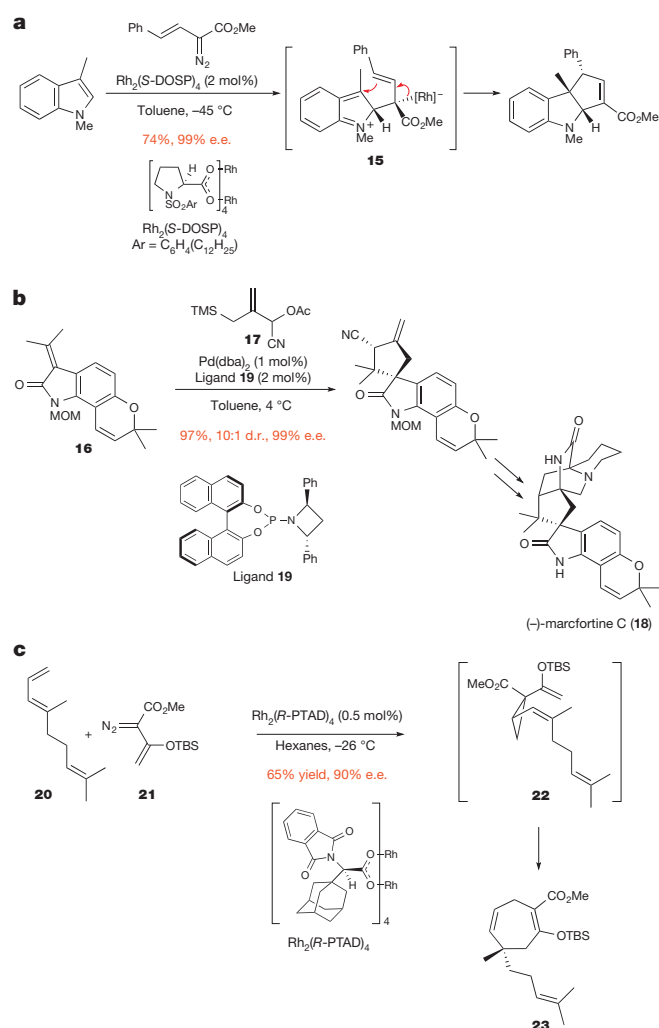
Figure 3 | Examples of other catalytic enantioselective cycloaddition reactions used to prepare products containing quaternary stereocentres.
a, The synthesis of a cyclopentene-fused indoline by a formal [3+2]-cycloaddition of 1,3-dimethylindole and a vinyl diazoester using a rhodium catalyst. This reaction is suggested to take place in a stepwise fashion via dipolar intermediate **15**[14]. **b**, The [3+2]-cycloaddition of a Pd-trimethylenemethane intermediate generated from allylic acetate **17** to form a tetracyclic intermediate in the total synthesis of (−)-marcfortine C[16]. MOM, methoxymethyl; TMS, trimethylsilyl; Ac, acetyl; dba, dibenzylideneacetone. **c**, Enantioselective synthesis of 1,4-cycloheptadiene **23** from triene **20** and vinyl diazoester **21**. The first step in this sequence is Rh-catalysed cyclopropanation of the terminal double bond of the acyclic triene to form divinyl cyclopropane **22**, which upon *in situ* Cope rearrangement generates **23** and its quaternary stereocentre. Product **23** was employed in the total synthesis of the diterpenoid (−)-5-*epi*-vibsanin E[21]. TBS, *tert*-butyldimethylsilyl.

formed from $SbCl_5$ and (*R*)-*o,o'*-dichloro-BINOL were the central steps[23,24]. Generally one equivalent of the complex was employed, although with structurally simpler substrates substoichiometric amounts could be employed (Fig. 4a).

The use of transition-metal catalysts has been more successful in achieving good catalytic efficiency in enantioselective polyene cyclizations. Especially promising are iridium-catalysed polyene cyclizations of allylic alcohol precursors developed by the Carreira group[25]. A variety of functionalized decalins containing angular substituents can be obtained in this way in useful yields and high enantioselectivity. For example, the transformation of triene allylic alcohol **24** to decalin **25** in 73% yield and 96% e.e. was the central step of a short total synthesis of (+)-asperolide C (**26**) (Fig. 4b)[26]. Toste and co-workers have described several gold-catalysed cyclizations that construct polycyclic products containing quaternary stereocentres, such as the dienyne polycyclization to form tetracyclic product

28 using ligand **29** (Fig. 4c)[27]. In addition, this group reported gold- and palladium-catalysed cyclizations of silyloxyenynes[28] and a palladium-catalysed variant of the Conia-ene reaction[29] to access functionalized cyclopentenes containing quaternary stereocentres. Rhodium catalysis has also been applied to the cyclization of dienynes to construct bicyclic, spirocyclic, and fused products depending upon the nature of the substrate. For example, the cyclization of acyclic dienyne **30** with a Rh/tol-BINAP catalyst led to the formation of bridged azatricyclic product **32** in 88% yield and 99% e.e.[30] (tol-BINAP, 2,2′-bis(di-*p*-tolylphosphino)-1,1′-binaphthalene). This reaction presumably takes place via metallacyclic intermediate **31**, which undergoes alkene insertion and reductive elimination to furnish product **32** (Fig. 4d).

The generation of chiral electrophiles to initiate polyene cyclizations using organic catalysts has also been developed in recent years. For example, the Jacobsen group reported the use of hydrogen-bonding thiourea catalysts to generate chiral *N*-acyliminium ion initiators of polyene cyclizations[31]. A novel approach to the catalytic enantioselective cyclization of acyclic polyenes was reported by the MacMillan group in which iminium ion activation and Cu(II)-promoted single-electron oxidation were combined to promote polycyclizations of radical-cation intermediates, as illustrated in the cyclization of polyene **33** (Fig. 4e)[32]. Hexacyclic product **35** is produced stereoselectively in 63% yield and 93% e.e. in a remarkable pentacyclization that begins with the generation of radical-cation intermediate **34**. The nitrile substituents were incorporated to favour 6-endo cyclizations of the radical intermediates, a requirement that is likely to limit the utility of this method for construction of natural terpenoids and steroids.

Transition–metal–catalysed insertions

In our earlier review of catalytic enantioselective synthesis of quaternary stereocentres, intramolecular Heck reactions were suggested to have the broadest demonstrated scope[5]. This method continues to be important. The total synthesis of (+)-minfiensine (**39**) reported by Overman and co-workers provides one recent illustration (Fig. 5a)[33]. In the Heck cyclization of dienyl triflate **36**, the use of the phosphinooxazoline (PHOX) ligand **40** was critical in achieving both high stereoinduction and preventing isomerization of the 1,4-diene product **37** to the conjugated 1,3-diene; avoiding double-bond migration was essential in allowing the second aza-cyclic ring of tetracyclic intermediate **38** to be generated upon exposure of the crude Heck product **37** to excess trifluoroacetic acid (TFA).

A variety of additional transition-metal-catalysed cyclization reactions have been developed recently for constructing polycyclic molecules containing quaternary stereocentres. Enantioselective nickel-catalysed intramolecular arylcyanation reactions disclosed by the groups of Jacobsen[34] and Nakao[35] are notable examples. The synthesis of indane **41** illustrates this transformation (Fig. 5b). An attractive feature of these isomerization reactions is the avoidance of the waste that would be generated in more conventional Heck-type cyclizations of related halide or triflate substrates. In another approach, Buchwald and co-workers reported the construction of quaternary stereocentres by palladium-catalysed cyclization/dearomatization of naphthalene derivatives. For example, tetracyclic amine **43** was obtained in high yield and enantioselectivity from bromodiarylamine precursor **42** using a catalyst generated from Pd(dba)$_2$ and binaphthyl ligand **44** (Fig. 5c)[36]. (Here dba is dibenzylideneacetone.) Dong and co-workers recently reported intramolecular carboacylation reactions of alkene-tethered benzocyclobutenones that construct various ring systems containing quaternary stereocentres, such as the formation of oxatricyclic ketone **47** from precursor **45** using a rhodium catalyst containing SEGPHOS ligand **48** (Fig. 5d)[37]. This transformation is suggested to proceed by initial formation of metallacyclic intermediate **46**.

Apparent in our discussion to this point is the prevalence of intramolecular reactions that construct quaternary stereocentres during ring formation. In this context, a recent report from the Sigman group is particularly important—the formation of aryl-containing quaternary stereocentres in high enantioselectivity by bimolecular Heck-type reactions of arylboronic acids and acyclic trisubstituted alkenes containing alcohol

a

b

Ligand 27

(+)-asperolide C (26)

c

Ligand 29

d

e
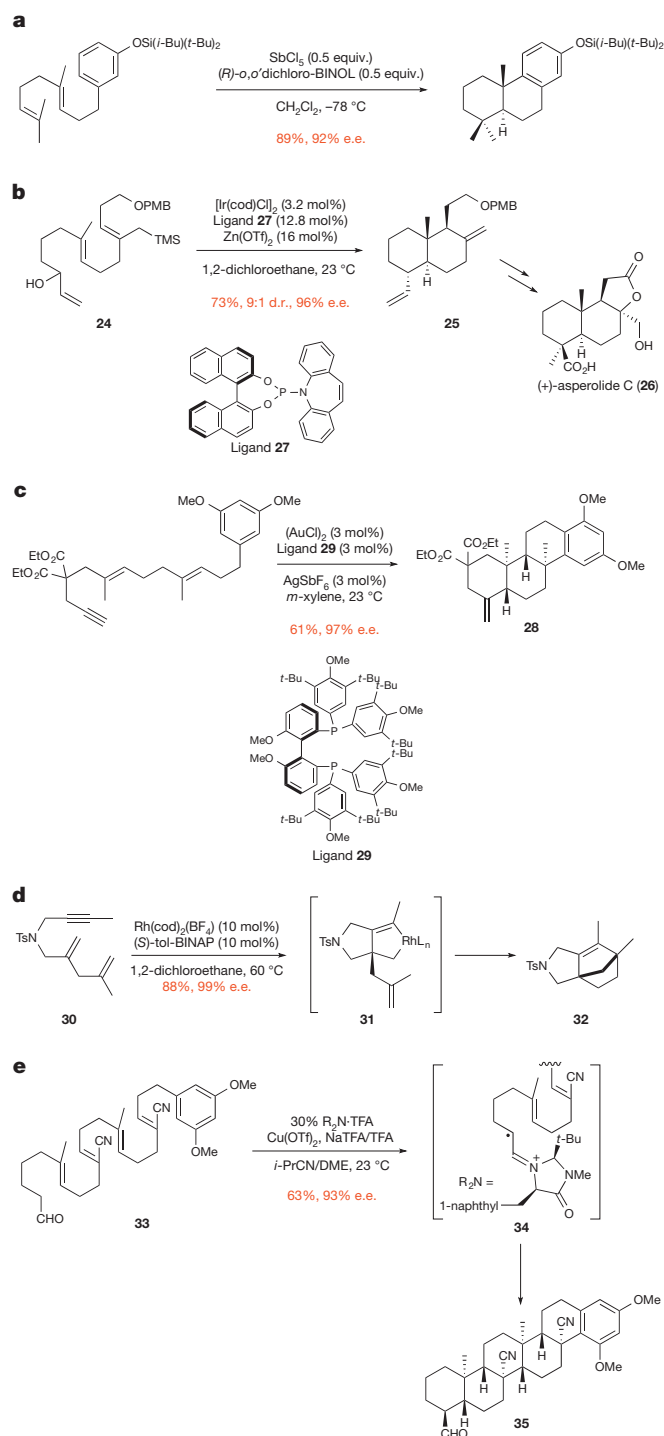
Figure 4 | Catalytic enantioselective polyene cyclizations to construct polycyclic products having quaternary stereocentres. a, The use of a protic acid catalyst for the cyclization of an aryl diene to form two rings and one quaternary stereocentre[24]. *i*-Bu, isobutyl; BINOL, 1,1′-bi-2-naphthol. **b**, The iridium-catalysed cyclization of a triene alcohol to construct the *trans*-decalin core **25** of the labdane diterpenoid (+)-asperolide C (**26**). The first step in this cascade cyclization is the generation of a η^3-allyliridium cation from the allylic alcohol fragment of **24**[26]. cod, 1,5-cyclooctadiene. **c**, The gold-catalysed cyclization of an aryl dienyne to form three rings and two quaternary stereocentres of tetracyclic product **28**[27]. **d**, The rhodium-catalysed cyclization of dienyne **30** to form bridged azatricyclic product **32**. This reaction is suggested to take place via metallacyclic intermediate **31**, which undergoes alkene insertion and reductive elimination to furnish product **32**[30]. Ts, *p*-toluenesulfonyl; tol-BINAP, 2,2′-bis(di-*p*-tolylphosphino)-1,1′-binaphthalene; L, ligand. **e**, The cyclization of tetraene aldehyde **33** in the presence of an imidazolone catalyst and a Cu(II) oxidant to form five rings and four quaternary stereocentres of hexacyclic product **35**. This novel reaction is suggested to proceed by single-electron oxidation of the initially formed iminium ion intermediate to generate **34**, which undergoes a series of 6-endo radical cyclizations to eventually give product **35**. The nitrile substituents are incorporated to disfavour 5-exo cyclizations in the formation of the second and fourth rings[32]. TFA, trifluoroacetic acid; NaTFA, sodium trifluoroacetate; *i*-Pr, isopropyl; DME, 1,2-dimethoxyethane.

progressed significantly over the past decade. Organocatalytic processes emerged to achieve such transformations, as well as numerous organometallic methods. Of particular note, high enantioselectivities can now be realized in copper-catalysed additions of various organometallic nucleophiles to prochiral Michael acceptors.

Ten years ago we noted that useful procedures for forming quaternary stereocentres by copper-catalysed additions of carbon nucleophiles to prochiral β,β-disubstituted enones and related electrophiles were notably absent[5]. This void is rapidly being filled, as many enantioselective copper-catalysed 1,4-addition reactions have now been reported that proceed with high enantioselectivities. Some of the more important of these methods are illustrated in Fig. 6 for conjugate additions to 3-methyl-2-cyclohexen-1-one (**54**). Copper-catalysed additions of various alkyl-[39], alkenyl-[40] and arylaluminium[41] compounds to cyclic enones in the presence of phosphoramidite ligands such as **58** have been described by Alexakis and co-workers (for example, **54**→**55**, Fig. 6a). Hoveyda and co-workers reported the use of Cu/Ag–NHC catalyst for the conjugate addition of alkyl- and arylaluminium intermediates to conjugated enones (for example, the synthesis of **56**, Fig. 6b)[42], as well as the addition of silicon-containing vinylaluminium intermediates[43]. The addition of arylzinc reagents can also be accomplished using a structurally related Cu/Ag–NHC catalyst generated from silver complex **62**, as exemplified in the formation of *ent*-**55** (Fig. 6c)[44]. Catalytic enantioselective conjugate addition reactions of other organometallic intermediates have also been reported. Notable examples include enantioselective copper-catalysed additions of Grignard reagents described by Alexakis and co-workers, as illustrated in the conversion of **54**→**57** (Fig. 6d)[45], and an alkene hydrozirconation/conjugate addition sequence reported by Fletcher and co-workers to construct cyclohexanone **58** using a copper catalyst containing phosphoramidite ligand **63** (Fig. 6e)[46].

Good success has also been realized in conjugate addition reactions that form quaternary stereocentres using rhodium and palladium catalysts. For example, Hayashi described the use of Rh(I) complexes of chiral dienes to catalyse the addition of arylboronic acids or tetraaryl boronates to maleimides[47] and enones[48]. Exemplary is the synthesis of cyclohexanone **55** in this fashion in 85% yield and 98% e.e. Rhodium-catalysed additions of arylaluminium reagents to β-substituted cyclic enones were also reported by the Alexakis group, including the synthesis of **55** in 71% yield and 98% e.e. using an Rh-BINAP catalyst[49]. In addition, Stoltz and co-workers reported palladium-catalysed variants of the 1,4-addition of boronic acids to enones for the enantioselective formation of chiral 3,3-disubstituted cyclohexanones[50].

As prochiral β,β-disubstituted α,β-unsaturated carbonyl compounds can be constructed in many ways and are common intermediates in retrosynthetic analysis, the recent development of versatile catalytic methods

substituents[38]. The enantioselective synthesis of ketone **52** from unsaturated alcohol **49** using a catalyst formed from Pd(CH$_3$CN)$_2$(OTs)$_2$ and diamine ligand **53** is exemplary (Fig. 5e). This transformation is suggested to proceed by initial enantioselective carbopalladation of the alkene to form intermediate **50**, followed by sequential β-hydride eliminations/migratory insertions along the alkyl chain to eventually yield alkene complex **51** and then ketone product **52**. Considerable variation of the substituents on the alkene is tolerated, and depending upon the starting alcohol, either ketones or aldehydes containing remote quaternary stereocentres can be formed in high enantioselectivity in this way.

Coupling of chiral carbon nucleophiles

The enantioselective formation of quaternary stereocentres by the coupling of chiral carbon nucleophiles with achiral carbon electrophiles has
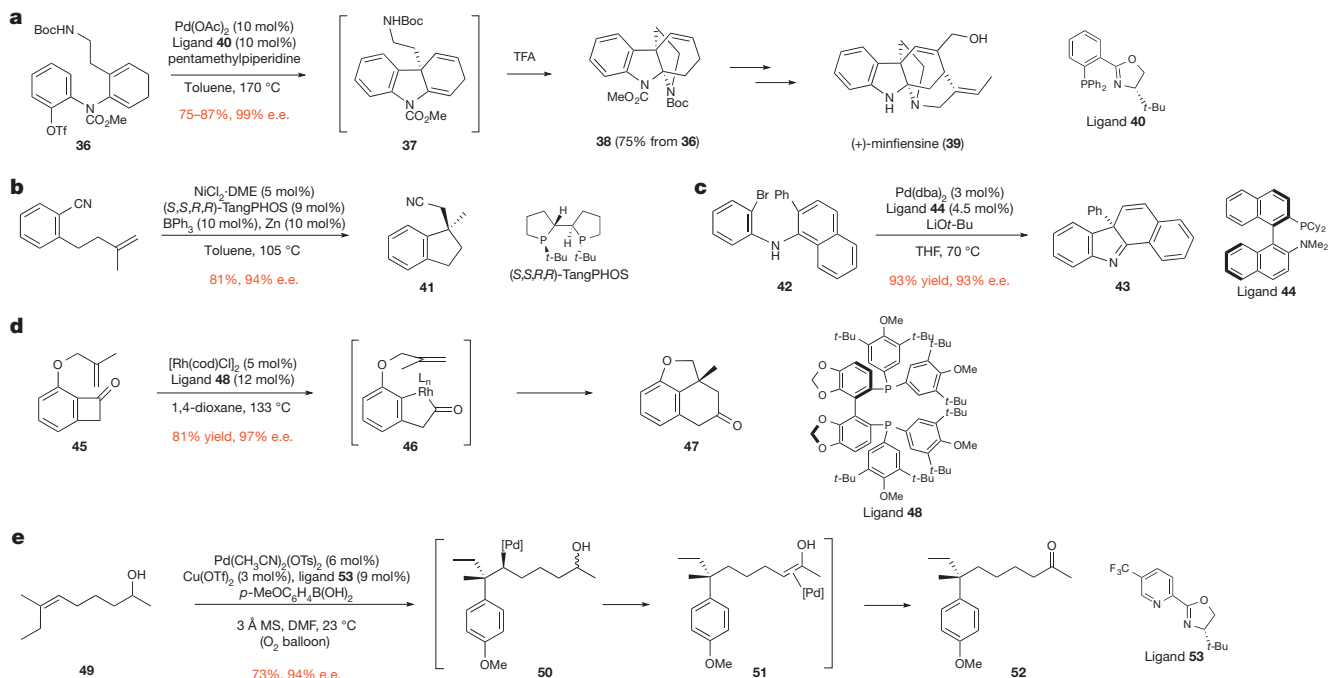
Figure 5 | Transition metal-catalysed insertion reactions that form quaternary stereocentres. a, The enantioselective intramolecular Heck cyclization of dienyl triflate **36** to form 1,4-diene intermediate **37**, which upon exposure to excess trifluoroacetic acid provided tetracyclic product **38** en route to the indole alkaloid (+)-minfiensine (**39**). The use of PHOX ligand **40** was critical in achieving both high stereoinduction and preventing isomerization of the initially formed product **37** to the conjugated 1,3-diene regioisomer[33]. **b**, The intramolecular nickel-catalysed arylcyanation of a tethered double bond to form indane **41**[34]. **c**, The palladium-catalysed cyclization/dearomatization of aryl(naphthyl)amine **42** to form tetracyclic product **43**. This reaction is suggested to occur via a six-membered palladacyclic intermediate that undergoes reductive elimination to form generate product **43**[36]. THF, tetrahydrofuran; Cy, cyclohexyl. **d**, The

rhodium-catalysed conversion of alkenyl benzocyclobutanone **45** to tricyclic ether **47**. This transformation is believed to occur by initial insertion of rhodium into the C–C bond to form acylrhodium intermediate **46**, which in the enantiodetermining step undergoes intramolecular carboacylation of the tethered alkene to form product **47**[37]. **e**, The bimolecular Heck-type addition of an arylboronic acid to the trisubstituted double bond of **49** to form ketone product **52**. This rare example of a bimolecular alkene insertion to form a quaternary stereocentre is suggested to occur by initial enantioselective carbopalladation of the alkene to generate intermediate **50**, which undergoes sequential β-hydride eliminations/migratory insertions along the alkyl chain to form alkene complex **51** and then the ketone product[38]. DMF, N,N-dimethylformamide.

to transform these intermediates into products containing new quaternary stereocentres is certain to find broad application. The transformations depicted in Fig. 6f, g, which were crucial steps in enantioselective total syntheses of (+)-taxa-4(5),11(12)-dien-2-one[51] and clavirolide C[52], are two recent examples.

When we discussed this approach for constructing quaternary stereocentres in our earlier review[5], organic catalysts—typically phase-transfer catalysts—had been employed with considerable success to join enolate intermediates with carbon electrophiles. The notable utility of cinchona alkaloid derivatives in such constructions has been further illustrated by Deng and co-workers in catalytic enantioselective additions of 1,3-dicarbonyl and related compounds to nitroalkenes and α,β-unsaturated ketones[53], and by Jørgensen in similar additions to allenic esters and ketones and for enantioselective alkynylations of 1,3-dicarbonyl compounds[54]. In addition, the use of enamine catalysis in the enantioselective construction of quaternary stereocentres from α-branched aldehydes—a reaction with broad potential utility for introducing quaternary stereocentres in a diversity of molecules[55]—was reported first by Barbas in 2004[56]. Organocatalysis has also been used to construct quaternary stereocentres by enantioselective intramolecular Stetter reactions of aromatic or aliphatic aldehydes (for example, **64**→**65**) using triazolium catalysts such as **66** (Fig. 7a)[57]. Other promising methods reported recently to exploit catalytically generated nucleophiles in the construction of quaternary stereocentres include the enantioselective insertion of diazoesters into the carbon–carbon bond of aryl aldehydes using an oxazaborolidinium catalyst[58], and the enantioselective alkylation of acyclic tributyltin enolates in the presence of a Cr(salen) catalyst[59]. (Here salen is N,N′-ethylenebis(salicylimine) 2,2′-ethylenebis(nitrilomethylidene)diphenoxide.)

A variety of enantioselective allylic substitution reactions have been reported in recent years that provide many opportunities for incorporating quaternary stereocentres in complex molecules. For example, a procedure developed by the Carreira group utilizes an iridium-cinchona alkaloid derivative dual-catalyst for the allylation of aldehydes. As exemplified in Fig. 7b, 3,3-disubstituted indoline **69** was constructed in this way with excellent enantio- and diastereoselectivity from indoline aldehyde **67** and allylic alcohol **68**[60]. Of most significance, catalytic enantioselective allylic alkylation reactions now allow quaternary stereocentres to be incorporated into many acyclic molecules or acyclic molecular fragments. The Hoveyda group has pioneered this area by introducing a variety of enantioselective copper-catalysed allylic substitution reactions[61–63]. In particular, this group has shown that a diverse array of carbon nucleophiles—such as dialkylzinc, vinylboron, vinylaluminium, and alkynylaluminium reagents—can be employed in allylic substitution reactions that form new quaternary stereocentres. The efficient and highly enantioselective alkylation of allylic phosphate **71** with an ester-containing vinylboron nucleophile to form product **72** in the presence of a copper-NHC catalyst is exemplary (Fig. 7c)[61]. As a final example, the enantio- and anti-diastereoselective allylic coupling of benzyl alcohol **74** with vinyl epoxide **75** to yield 1,3-diol **76** using Ir catalyst **77** reported by Krische and co-workers even allows a benzyl alcohol to be employed as the pro-electrophile in the construction of quaternary stereocentres (Fig. 7d)[64]. This reaction, which results in appending a 1-(hydroxymethyl)-1-methylallyl unit to the alcohol fragment, should find use in the synthesis of terpenoid natural products that incorporate this (hydroxy)prenyl motif. A notable feature of redox-triggered couplings of this type pioneered by the Krische group is the absence of stoichiometric by-products.
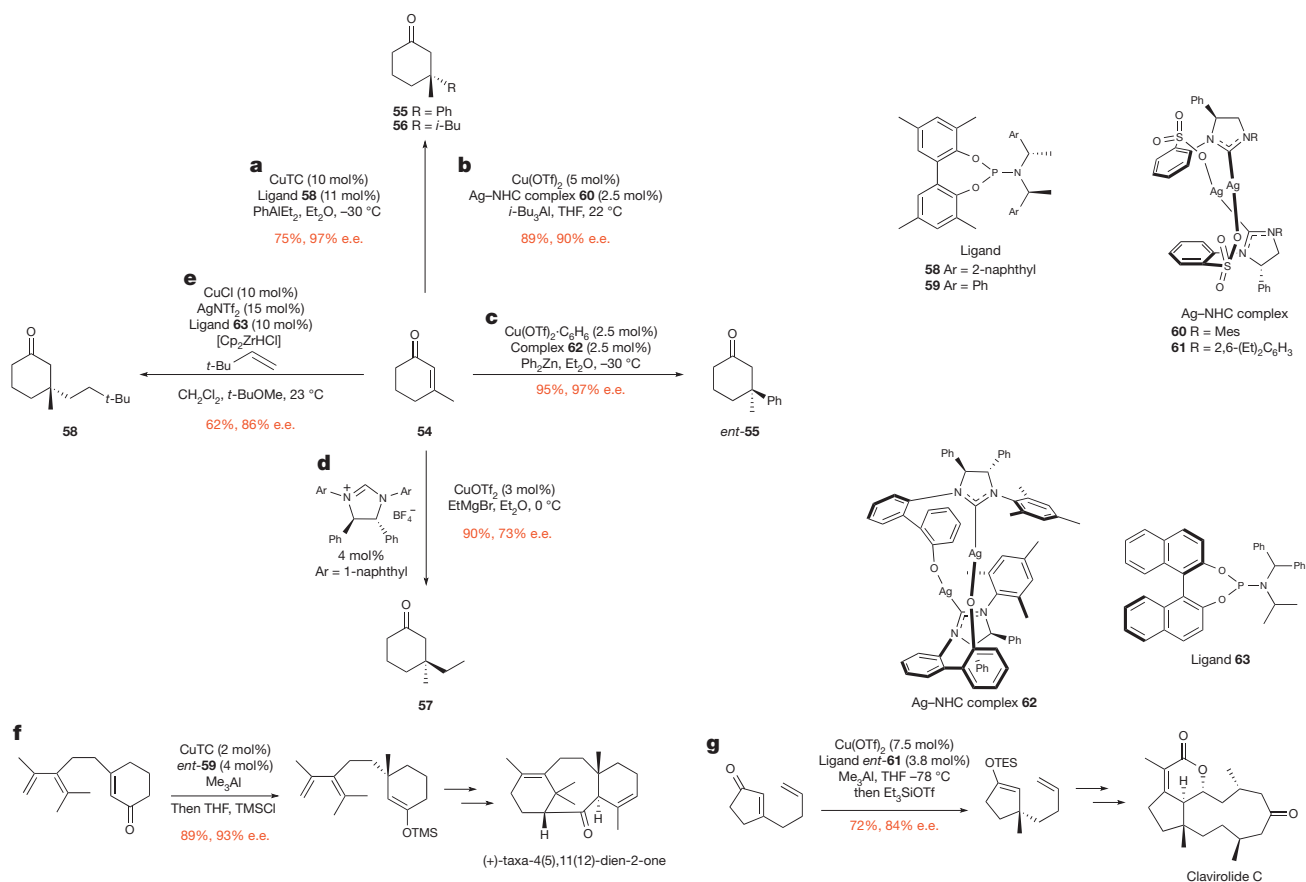
Figure 6 | Enantioselective copper-catalysed conjugate additions to construct quaternary stereocentres. a–e, Cu-catalysed conjugate additions to 3-methyl-2-cyclohexen-1-one (**54**) that form new quaternary stereocentres. **a,** The addition of an arylaluminium compound to **54** to form cyclohexanone **55**[40]. CuTC, copper(I) thiophene-2-carboxylate. **b,** The addition of a trialkylaluminium compound to **54** to form cyclohexanone **56**[42]. NHC, N-heterocyclic carbene. **c,** The addition of an arylzinc compound to **54** to form the enantiomer of cyclohexanone **55**[44]. **d,** The addition of an alkyl Grignard reagent to **54** to form 3,3-dialkylcyclohexanone **57**[45]. **e,** The addition of an alkylzirconium intermediate generated by hydrozirconation of 3,3-dimethyl-1-butene to **54** to form 3,3-dialkylcyclohexanone **58**[46]. Cp, cyclopentadienyl. **f, g,** Use of two of these methods to form methyl-containing quaternary stereocentres in syntheses of a potential taxane terpenoid precursor and a dolabellane diterpenoid. **f,** The enantioselective copper-catalysed conjugate addition/enolate trapping to introduce a quaternary methyl group in the construction of a taxadienone[51]. **g,** The enantioselective copper-catalysed conjugate addition/enolate trapping to introduce a quaternary methyl group in the total synthesis of clavirolide C[52]. TES, triethylsilyl.
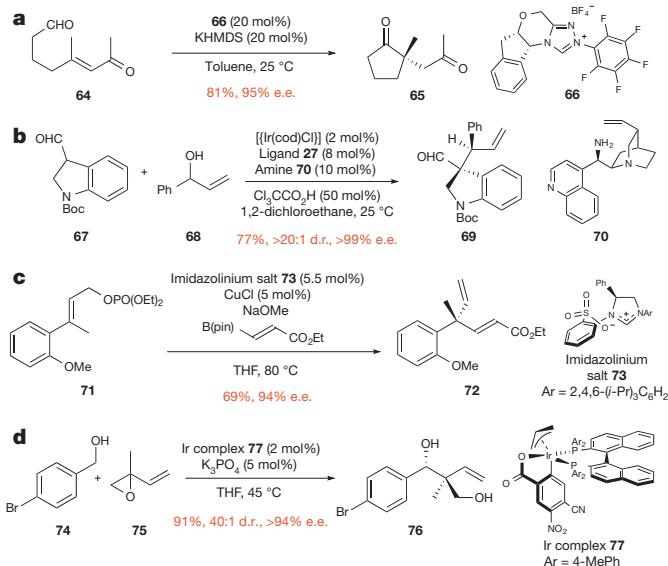
Figure 7 | Use of the enantioselective intramolecular Stetter reaction and allylic alkylation reactions to construct quaternary stereocentres. a, The intramolecular Stetter reaction of enone aldehyde **64** catalysed by the carbene generated from triazolium salt **66** to form 2,2-disubstituted cyclopentanone **65**[57]. KHMDS, potassium bis(trimethylsilyl)amide. **b,** The α-allylation of indoline aldehyde **67** with allylic alcohol **68** using the dual activity of iridium and amine catalysts. This reaction constructs the quaternary and adjacent secondary stereocentre of product **69**[60]. **c,** The formation of 4,4-disubstituted 2,5-hexadienoic ester **72** by allylic displacement of phosphate triester **71**. In this reaction, an alkenylcopper carbene complex is generated from a vinylboronate precursor[61]. pin, pinacolato. **d,** The anti-diastereoselective coupling of benzyl alcohol **74** with vinyl epoxide **75** using an iridium catalyst to give the product **76** of carbonyl tert-(hydroxy)prenylation. This reaction proceeds by the coupling of aldehyde and (E)-σ-allyliridium intermediates respectively generated in situ from the alcohol and vinyl epoxide precursors by an iridium-catalysed redox process[64].

Coupling of chiral carbon electrophiles

Reactions of chiral carbon electrophiles with carbon nucleophiles encompass a range of transformations that can be used to form quaternary stereocentres in structurally complex molecules. During the past decade, palladium-catalysed enantioselective allylic alkylation reactions have been applied widely to achieve this aim, and a number of other promising methods employing transition metal or organic catalysts have been introduced.

The use of enantioselective palladium-catalysed allylic alkylation reactions to form quaternary stereocentres adjacent to ketone carbonyl groups was initially reported by the Stoltz[65] and Trost groups[66]. Since these initial reports, this method has been featured in several natural product total syntheses[4]. For example, a variety of chiral 3,3-disubstituted oxindoles have been prepared in this fashion with good enantioselectivity[67,68], as exemplified by the enantioselective and regioselective prenylation (78→79) used in the synthesis of *ent*-flustramines A (80) and B (Fig. 8a)[69]. In a strategically incisive example, Stoltz and co-workers employed an enantioselective double allylation of racemic bis-β-ketoester 82 to form C_2-symmetric diketone 83 en route to (−)-cyanthiwigin F (84) (Fig. 8b)[70]. Other significant recent developments in this area include the use of a vinyl epoxide as a coupling partner in the total syntheses of (−)-biyouyanagin A and hyperolactone C[71], and the application of molybdenum[72] and iridium[73] catalysts in enantioselective allylic alkylation reactions. In addition, the enantioselective C-3 allylation of an indole derivative using allyl alcohol in combination with a trialkylborane as the alkylating reagent was featured in a synthesis of (−)-esermethole[74].

Organocatalytic reactions can also be employed to generate chiral electrophiles for constructing quaternary stereocentres. Particularly well developed is the use of catalytic enantioselective Steglich rearrangements. Fu[75] and Vedejs[76] developed chiral-enantiopure variants of 4-(dimethylamino)pyridine (DMAP) to accomplish enantioselective rearrangements of enoxycarbonate derivatives, including those derived from oxindoles and furanones. The Fu group also described related transformations involving the acylation of silyl ketene imines and employed this method as the central step in a synthesis of (S)-verapamil[77]. In a concise second-generation total synthesis of (+)-gliocladin C (87), Overman and co-workers exploited the planar-chiral DMAP variant 88[75] to catalyse the enantioselective Steglich rearrangement of enoxycarbonate 85 to yield oxindole 86 (Fig. 9a)[78]. In this study, the practicality of Fu's method was highlighted by the formation of 86 in 96% yield and 96% e.e. on multigram scales. In a quite different approach to generating chiral carbon electrophiles, iminium activation developed by the MacMillan group has been used for the enantioselective construction of 3a-substituted pyrrolidinoindolines and featured in the synthesis of (−)-flustramine B[79].

Although their scope is less well defined at this point than enantioselective palladium-catalysed allylation reactions or Steglich rearrangements, enantioselective transition-metal-catalysed arylations, vinylations, and alkylations of prochiral nucleophiles have been described recently for the enantioselective construction of quaternary stereocentres. One example is the enantioselective copper-catalysed indole arylation/cyclisation
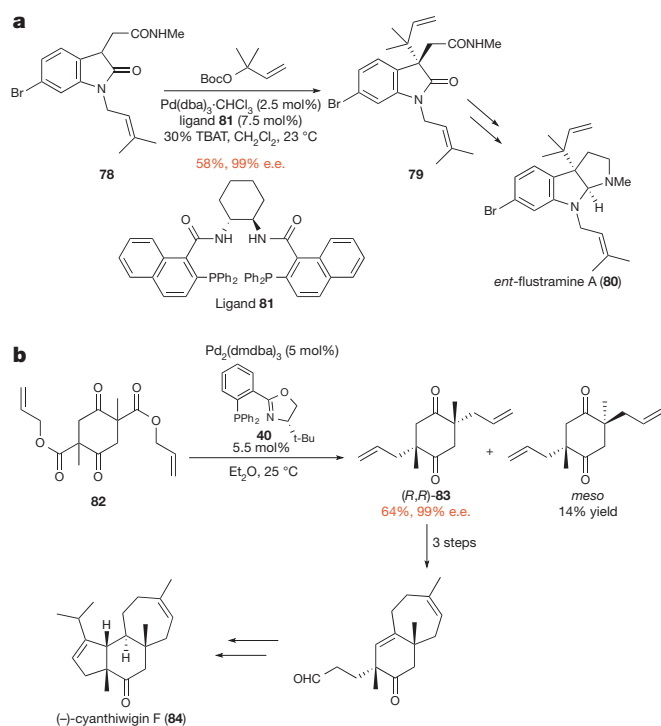
Figure 8 | Use of palladium-catalysed asymmetric allylic alkylation reactions for constructing quaternary centres in alkaloid and terpenoid natural products. a, The regioselective prenylation of oxindole 78 upon base-promoted reaction with the η³-allylpalladium electrophile generated from a prenyl carbonate to form 79. This product was a late-stage intermediate in the enantioselective total synthesis of *ent*-flustramine A (80)[69]. TBAT, tetrabutylammonium difluorotriphenylsilicate. **b**, The *syn*-diastereoselective diallylation of β-ketoester 82 (a mixture of racemic diastereomers) to give (R,R)-83, a pivotal intermediate in the enantioselective total synthesis of (−)-cyanthiwigin F[70]. dmdba, bis(3,5-dimethoxybenzylidene)acetone.
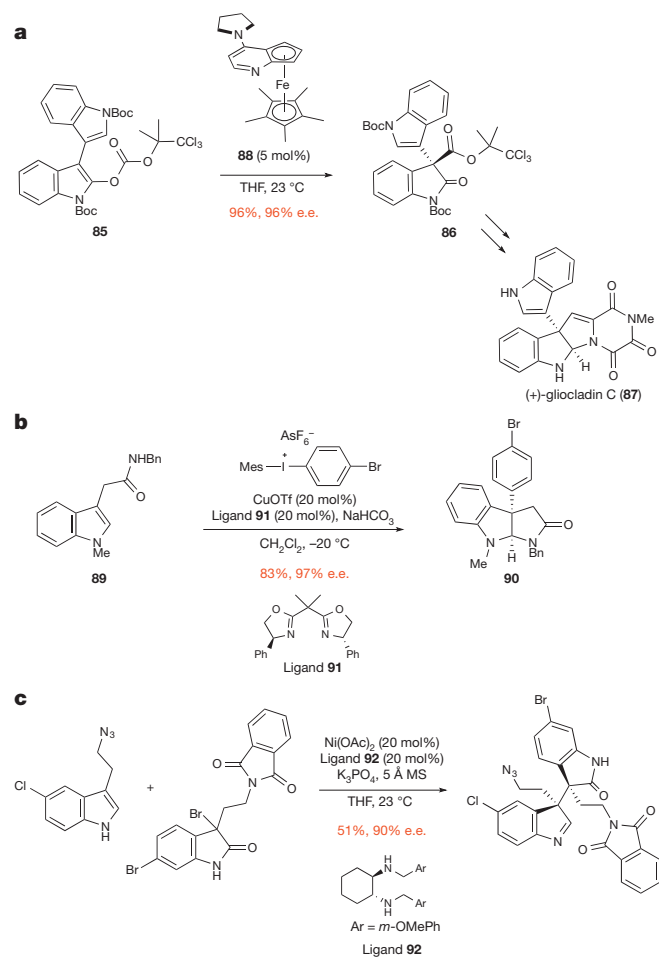
Figure 9 | Miscellaneous methods involving the union of a catalytically generated chiral carbon electrophile with a carbon nucleophile. a, The Steglich rearrangement of indole carbonate 85 in the presence of Fu's planar-chiral catalyst 88 to give 3,3-disubstituted oxindole 86 en route to (+)-gliocladin C[78]. **b**, The copper-catalysed β-arylation of indole 89 and concomitant cyclisation to form 3a-arylpyrrolidinoindolinone 90[80]. Bn, benzyl; Mes, 1,3,5-trimethylbenzene. **c**, The Ni-catalysed coupling of an indole with a 3-bromooxindole en route to (+)-peroporamidine. This reaction sets the two contiguous quaternary stereocentres of (+)-peroporamidine[85]. OAc, acetoxy.

sequence reported by the MacMillan group[80]. In this transformation, tryptophan amides undergo efficient indole arylation in the presence of a diaryliodonium salt, CuOTf and enantiopure bisoxazoline ligand 91, followed by intramolecular trapping of the pendant amide to form 3a-arylpyrrolidinoindolines (89→90) (Fig. 9b). Other notable examples of copper-catalysed arylation and vinylation reactions used to construct quaternary stereocentres are the copper-catalysed arylations of prochiral β-ketoesters with 2-iodotrifluoroacetanilides described by Ma and co-workers[81], and the palladium-catalysed enantioselective α-arylations of α-branched aldehydes[82] and C-3 arylations or vinylations of oxindoles reported by the Buchwald group[83]. Catalytic enantioselective alkylations of prochiral nucleophiles can be achieved as well. A double alkylation of a 3,3′-dioxindole with nitroethylene was reported by the Shibasaki group en route to (+)-chimonanthine, (+)-folicanthine and (−)-calycanthine[84]. In a mechanistically intriguing variant, catalytic enantioselective alkylations of 3-bromooxindoles with 3-substituted indoles were reported by the Wang group for the construction of vicinal quaternary stereocentres using a catalyst formed from Ni(OAc)₂ and diamine ligand 92. This step in the total synthesis of (+)-perophoramidine is illustrated in Fig. 9c[85]. This reaction is suggested to occur by loss of HBr from the 3-bromooxindole to generate an electrophilic indol-2one intermediate, which couples with the indole nucleophile. How the nickel-diamine catalyst organizes this coupling to achieve high enantio- and diastereoselectivity is unclear at present.

Desymmetrization reactions

In principle, any catalytic enantioselective reaction could be employed to construct a product containing a quaternary stereocentre by desymmetrization of an appropriately constituted prochiral precursor. Since our earlier review[5], numerous additional examples of using group-selective catalytic enantioselective reactions for this purpose have been described. For instance, in their synthesis of (+)-quebrachamine (95), Hoveyda and Schrock reported the use of a chiral molybdenum metathesis catalyst to fashion the tetrahydropyridine ring of intermediate 94 from triene precursor 93 in excellent yield and enantioselectivity (Fig. 10a)[86]. The Hoveyda group has also reported the use of enantioselective ring-opening/cross-metathesis to construct acyclic products bearing quaternary centres[87]. In a quite different approach reported by the Toste group, cyclobutanones containing α-quaternary stereocentres can be prepared by enantioselective gold-catalysed ring expansion of prochiral allenylcyclopropanols (Fig. 10b)[88]. Applications of two recently developed rhodium-catalysed C–C-bond constructions for enantioselective desymmetrization are exemplified in Fig. 10c, d. In the first example, from the laboratory of Dong, intermolecular hydroacylation of the prochiral cyclopropene 96 with salicyaldehyde delivers the highly substituted cyclopropane product 97 in high yield and enantiomeric purity (Fig. 10c)[89]. The second example, from the Cramer group, illustrates the use of C–C bond activation in the efficient and enantioselective formation of bridged tricyclic ketone 99 from the prochiral cyclobutanone precursor 98 (Fig. 10d)[90]. In an example
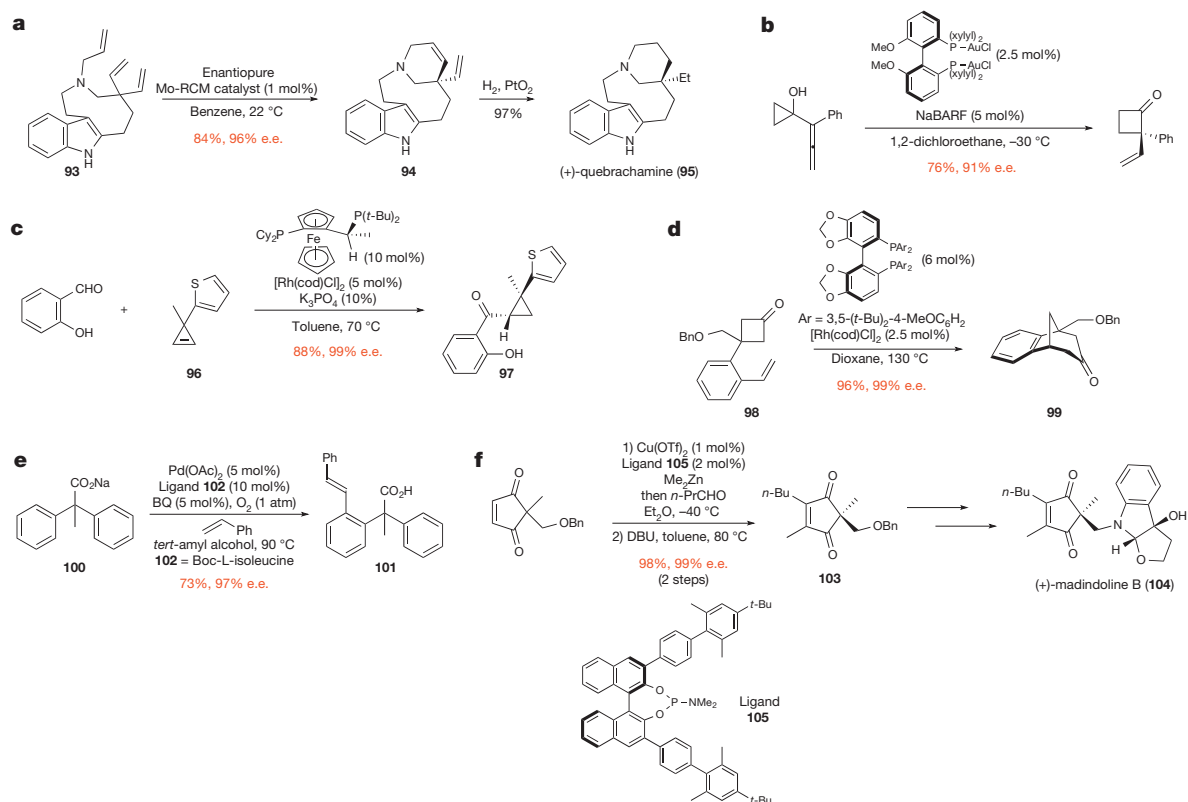
Figure 10 | Enantioselective desymmetrization reactions of precursors containing prochiral quaternary carbons. a, The ring-closing metathesis of triene 93 to give tetrahydropyridine 94 using a molybdenum catalyst. Catalytic hydrogenation of product 94 then completes a novel construction of (+)-quebrachamine (95)[86]. RCM, ring-closing metathesis. **b**, The gold-catalysed ring expansion of an allenylcyclopropanol to form (R)-2-ethenyl-2-phenylcyclobutanone[88]. xylyl, 3,5-dimethylphenyl; NaBARF, sodium tetrakis[3,5-bis(trifluoromethyl)phenyl]borate. **c**, The rhodium-catalysed hydroacylation of cyclopropene 96 with salicylaldehyde to form cyclopropane 97. Coordination of the phenolic oxygen of salicylaldehyde and the ring strain of the cyclopropene promotes this bimolecular hydroacylation reaction. The observed diastereoselectivity is suggested to result from rhodium-hydride

insertion and subsequent C–C bond reductive elimination taking place preferentially from the cyclopropene face opposite the larger substituent[89]. **d**, The enantiotopic rhodium-catalysed insertion into a C–C bond of cyclobutanone 98, followed by intramolecular insertion of the rhodium-acyl intermediate to give bridged-tricyclic ketone 99[90]. **e**, The palladium(II)-catalysed enantiotopic C–H activation of sodium diphenylacetate 100 templated by the carboxylate group, followed by bimolecular Heck coupling with styrene to give product 101[91]. BQ, benzoquinone. **f**, The desymmetrization of a prochiral 1,4-cyclopentenone by copper-catalysed conjugate addition of a methyl group to give chiral product 103 was the key step in the total synthesis of (+)-madindoline B[92] (104). DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene.

exploiting enantioselective C–H activation, the Yu group reported palladium(II)-catalysed group-selective functionalizations of diphenylacetic acid derivatives[91]. Using a palladium catalyst containing protected amino acid ligands, various diphenylacetic derivatives underwent selective alkenylation with acrylates or styrenes, exemplified by the conversion of **100**→**101** (Fig. 10e). In a final example, a number of prochiral cyclopentene-1,3-diones have been desymmetrized by copper-catalysed enantioselective additions of dialkylzinc or organoaluminium reagents[92]. The use of a phosphoramidite ligand such as **105** proved optimal in this method, as illustrated in the enantioselective synthesis of cyclopentene-1,3-dione **103**, a key step in the synthesis of (+)-madindoline B (**104**) (Fig. 10f). Organocatalytic methods have also proven useful for constructing quaternary stereocentres by desymmetrization. Protic-acid catalysed vinylogous α-ketol rearrangements to yield spirocyclic diones[93], and the preparation of five-membered rings from 1,3-diketone precursors using chiral NHC-catalysts are two important examples[94].

Looking forward

The research highlighted in this brief survey shows that a variety of chemical transformations are now available to synthetic chemists for incorporating quaternary stereocentres in organic molecules with high enantioselectivity. When the catalytic transformations that are the focus of our analysis are combined with non-catalytic methods, a diversity of chemical transformations are now available to meet this formidable challenge. Nonetheless, the scope of the majority of the methods discussed in this Review is only partially defined, and limitations are certain to be uncovered. One area where the development of methods is still in its early stages is the introduction of quaternary stereocentres in acyclic molecules or acyclic molecular fragments[95]. Even in areas where substantial progress has been recorded recently in fashioning quaternary stereocentres in cyclic molecules—for example, by conjugate additions to cyclohexenones—enantioselectivities realized in identical reactions with cyclic enones of other ring sizes or acyclic enones can be inferior. It is instructive to note that almost all the methods exemplified in this Review involve the functionalization of π-bonds. With the intense attention currently being paid to the direct functionalization of Csp^3–H σ-bonds, we anticipate that catalytic C–H insertions will play a much larger role in the future in the enantioselective synthesis of quaternary stereocentres. For example, the scope of such transformations for desymmetrizing prochiral quaternary carbons is certain to expand[96], and new methods exploiting selective C–H functionalizations will probably be developed for transforming chiral tertiary carbons (enantiopure or racemic) and prochiral secondary carbons to new quaternary stereocentres. Finally, as nearly half of the transformations we exemplified involve the use of catalysts containing rare and/or expensive metals, we note that the development of alternative catalytic methods based on readily available and less expensive catalysts remains a critical future challenge in this area.

The methods now available for fashioning quaternary stereocentres enantioselectively remove much of the previous barrier to incorporating such functionality in organic molecules for use in medicine, agriculture and other areas where high-value organic molecules play an important role. One can already see this impact in the structure of a few small molecules currently undergoing clinical evaluation, such as anamorelin[97]. With several recent studies suggesting that drug candidates that contain a larger fraction of sp^3 carbons and chiral centres have a lower rate of attrition in the clinic[98], we anticipate seeing an ever increasing number of drug candidates containing quaternary stereocentres being designed, synthesized and evaluated.

1. Bartholow, M. Top 200 drugs of 2011. *Pharm. Times* 48–51 (10 July 2012).
2. Ding, H. X., Liu, K. K.-C., Sakya, S. M., Flick, A. C. & O'Donnell, C. J. Synthetic approaches to the 2011 new drugs. *Bioorg. Med. Chem.* **21,** 2795–2825 (2013).
3. Christoffers, J. & Baro, A. (eds) *Quaternary Stereocenters — Challenges and Solutions for Organic Synthesis* (Wiley-VCH, 2005).
4. Hong, A. Y. & Stoltz, B. M. The construction of all-carbon quaternary stereocenters by use of Pd-catalyzed asymmetric allylic alkylation reactions in total synthesis. *Eur. J. Org. Chem.* 2745–2759 (2013).
5. Douglas, C. J. & Overman, L. E. Catalytic asymmetric synthesis of all-carbon quaternary stereocenters. *Proc. Natl Acad. Sci. USA* **101,** 5363–5367 (2004).
6. Wilson, R. M., Jen, W. S. & MacMillan, D. W. C. Enantioselective organocatalytic intramolecular Diels–Alder reactions. The asymmetric synthesis of solanapyrone D. *J. Am. Chem. Soc.* **127,** 11616–11617 (2005).
7. Jones, S. B., Simmons, B., Mastracchio, A. & MacMillan, D. W. C. Collective synthesis of natural products by means of organocascade catalysis. *Nature* **475,** 183–188 (2011).
   A highlight of the use of iminium activation and cascade catalysis in the synthesis of alkaloid natural products.
8. Laforteza, B. N., Pickworth, M. & MacMillan, D. W. C. Enantioselective total synthesis of (–)-minovincine in nine chemical steps: an approach to ketone activation in cascade catalysis. *Angew. Chem. Int. Edn* **52,** 11269–11272 (2013).
9. Liu, Y., Nappi, M., Arceo, E., Vera, S. & Melchiorre, P. Asymmetric catalysis of Diels–Alder reactions with in situ generated heterocyclic ortho-quinodimethanes. *J. Am. Chem. Soc.* **133,** 15212–15218 (2011).
10. Tan, B., Hernández-Torres, G. & Barbas, C. F. III Highly efficient hydrogen-bonding catalysis of the Diels–Alder reaction of 3-vinylindoles and methyleneindolinones provides carbazolespirooxindole skeletons. *J. Am. Chem. Soc.* **133,** 12354–12357 (2011).
11. Shimizu, Y., Shi, S.-L., Usuda, H., Kanai, M. & Shibasaki, M. Catalytic asymmetric total synthesis of ent-hyperforin. *Angew. Chem. Int. Edn* **49,** 1103–1106 (2010).
12. Snyder, S. A. & Corey, E. J. Concise total syntheses of palominol, dolabellatrienone, β-araneosene, and isoedunol via an enantioselective Diels–Alder macrobicyclization. *J. Am. Chem. Soc.* **128,** 740–742 (2006).
13. Balskus, E. P. & Jacobsen, E. N. Asymmetric catalysis of the transannular Diels–Alder reaction. *Science* **317,** 1736–1740 (2007).
14. Lian, Y. & Davies, H. M. L. Rhodium-catalyzed [3 + 2] annulation of indoles. *J. Am. Chem. Soc.* **132,** 440–441 (2010).
15. Trost, B. M., Silverman, S. M. & Stambuli, J. P. Development of an asymmetric trimethylenemethane cycloaddition reaction: application in the enantioselective synthesis of highly substituted carbocycles. *J. Am. Chem. Soc.* **133,** 19483–19497 (2011).
16. Trost, B. M., Bringley, D. A., Zhang, T. & Cramer, N. Rapid access to spirocyclic oxindole alkaloids: application of the asymmetric palladium-catalyzed [3 + 2] trimethylenemethane cycloaddition. *J. Am. Chem. Soc.* **135,** 16720–16735 (2013).
   An informative discussion of development of the TMM reaction for the synthesis of spirocyclic oxindoles.
17. Ohmatsu, K., Imagawa, N. & Ooi, T. Ligand-enabled multiple absolute stereocontrol in metal-catalysed cycloaddition for construction of contiguous all-carbon quaternary stereocentres. *Nature Chem.* **6,** 47–51 (2014).
18. Chen, X.-H., Wei, Q., Luo, S.-W., Xiao, H. & Gong, L.-Z. Organocatalytic synthesis of spiro[pyrrolidin-3,3′-oxindoles] with high enantiopurity and structural diversity. *J. Am. Chem. Soc.* **131,** 13819–13825 (2009).
19. Martínez, A., Webber, M. J., Müller, S. & List, B. Versatile access to chiral indolines by catalytic asymmetric Fischer indolization. *Angew. Chem. Int. Edn* **52,** 9486–9490 (2013).
20. Pellissier, H. Recent developments in asymmetric cyclopropanation. *Tetrahedron* **64,** 7041–7095 (2008).
21. Schwartz, B. D., Denton, J. R., Lian, Y., Davies, H. M. L. & Williams, C. M. Asymmetric [4 + 3] cycloadditions between vinylcarbenoids and dienes: application to the total synthesis of the natural product (–)-5-epi-vibsanin E. *J. Am. Chem. Soc.* **131,** 8329–8332 (2009).
22. Ishibashi, H., Ishihara, K. & Yamamoto, H. A new artificial cyclase for polyprenoids: enantioselective total synthesis of (–)-chromazonarol, (+)-8-epi-puupehedione, and (–)-11′-deoxytaondiol methyl ether. *J. Am. Chem. Soc.* **126,** 11122–11123 (2004).
23. Surendra, K. & Corey, E. J. Highly enantioselective proton-initiated polycyclization of polyenes. *J. Am. Chem. Soc.* **134,** 11992–11994 (2012).
24. Surendra, K., Rajendar, G. & Corey, E. J. Useful catalytic enantioselective cationic double annulation reactions initiated at an internal π-bond: method and applications. *J. Am. Chem. Soc.* **136,** 642–645 (2014).
25. Schafroth, M. A., Sarlah, D., Krautwald, S. & Carreira, E. M. Iridium-catalyzed enantioselective polyene cyclization. *J. Am. Chem. Soc.* **134,** 20276–20278 (2012).
   A potentially broadly applicable approach for orchestrating enantioselective polyene cyclizations.
26. Jeker, O. F., Kravina, A. G. & Carreira, E. M. Total synthesis of (+)-asperolide C by iridium-catalyzed enantioselective polyene cyclization. *Angew. Chem. Int. Edn* **52,** 12166–12169 (2013).
27. Sethofer, S. G., Mayer, T. & Toste, F. D. Gold(I)-catalyzed enantioselective polycyclization reactions. *J. Am. Chem. Soc.* **132,** 8276–8277 (2010).
28. Brazeau, J.-F., Zhang, S., Colomer, I., Corkey, B. K. & Toste, F. D. Enantioselective cyclizations of silyloxyenynes catalyzed by cationic metal phosphine complexes. *J. Am. Chem. Soc.* **134,** 2742–2749 (2012).
29. Corkey, B. K. & Toste, F. D. Catalytic enantioselective Conia-ene reaction. *J. Am. Chem. Soc.* **127,** 17168–17169 (2005).
30. Shibata, T., Tahara, Y.-K., Tamura, K. & Endo, K. Enantioselective syntheses of various chiral multicyclic compounds with quaternary carbon stereocenters by catalytic intramolecular cycloaddition. *J. Am. Chem. Soc.* **130,** 3451–3457 (2008).
31. Knowles, R. R., Lin, S. & Jacobsen, E. N. Enantioselective thiourea-catalyzed cationic polycyclizations. *J. Am. Chem. Soc.* **132,** 5030–5032 (2010).

32. Rendler, S. & MacMillan, D. W. C. Enantioselective polyene cyclization via organo-SOMO catalysis. *J. Am. Chem. Soc.* **132,** 5027–5029 (2010).

33. Dounay, A. B., Humphreys, P. G., Overman, L. E. & Wrobleski, A. D. Total synthesis of the *strychnos* alkaloid (+)-minfiensine: tandem enantioselective intramolecular Heck–iminium ion cyclization. *J. Am. Chem. Soc.* **130,** 5368–5377 (2008).

34. Watson, M. P. & Jacobsen, E. N. Asymmetric intramolecular arylcyanation of unactivated olefins via C–CN bond activation. *J. Am. Chem. Soc.* **130,** 12594–12595 (2008).

35. Nakao, Y. et al. Intramolecular arylcyanation of alkenes catalyzed by nickel/AlMe$_2$Cl. *J. Am. Chem. Soc.* **130,** 12874–12875 (2008).

36. García-Fortanet, J., Kessler, F. & Buchwald, S. L. Palladium-catalyzed asymmetric dearomatization of naphthalene derivatives. *J. Am. Chem. Soc.* **131,** 6676–6677 (2009).

37. Xu, T., Ko, H. M., Savage, N. A. & Dong, G. Highly enantioselective Rh-catalyzed carboacylation of olefins: efficient syntheses of chiral poly-fused rings. *J. Am. Chem. Soc.* **134,** 20005–20008 (2012).

38. Mei, T.-S., Patel, H. H. & Sigman, M. S. Enantioselective construction of remote quaternary stereocentres. *Nature* **508,** 340–344 (2014).
A pioneering method for constructing remote stereocentres in acyclic substrates by enantioselective Heck reactions.

39. Vuagnoux-d'Augustin, M. & Alexakis, A. Copper-catalyzed asymmetric conjugate addition of trialkylaluminium reagents to trisubstituted enones: construction of chiral quaternary centers. *Chem. Eur. J.* **13,** 9647–9662 (2007).

40. Müller, D. & Alexakis, A. Formation of quaternary stereogenic centers by copper-catalyzed asymmetric conjugate addition reactions of alkenylaluminums to trisubstituted enones. *Chem. Eur. J.* **19,** 15226–15239 (2013).

41. Hawner, C., Li, K., Cirriez, V. & Alexakis, A. Copper-catalyzed asymmetric conjugate addition of aryl aluminum reagents to trisubstituted enones: construction of aryl-substituted quaternary centers. *Angew. Chem. Int. Edn* **47,** 8211–8214 (2008).

42. May, T. L., Brown, M. K. & Hoveyda, A. H. Enantioselective synthesis of all-carbon quaternary stereogenic centers by catalytic asymmetric conjugate additions of alkyl and aryl aluminum reagents to five-, six-, and seven-membered-ring β-substituted cyclic enones. *Angew. Chem. Int. Edn* **47,** 7358–7362 (2008).

43. May, T. L., Dabrowski, J. A. & Hoveyda, A. H. Formation of vinyl-, vinylhalide- or acyl-substituted quaternary carbon stereogenic centers through NHC–Cu-catalyzed enantioselective conjugate additions of Si-containing vinylaluminums to β-substituted cyclic enones. *J. Am. Chem. Soc.* **133,** 736–739 (2011).

44. Lee, K.-S., Brown, M. K., Hird, A. W. & Hoveyda, A. H. A practical method for enantioselective synthesis of all-carbon quaternary stereogenic centers through NHC–Cu-catalyzed conjugate additions of alkyl and arylzinc reagents to β-substituted cyclic enones. *J. Am. Chem. Soc.* **128,** 7182–7184 (2006).

45. Martin, D. et al. Copper-catalyzed asymmetric conjugate addition of Grignard reagents to trisubstituted enones. Construction of all-carbon quaternary chiral centers. *J. Am. Chem. Soc.* **128,** 8416–8417 (2006).

46. Sidera, M., Roth, P. M. C., Maksymowicz, R. M. & Fletcher, S. P. Formation of quaternary carbon centers by copper-catalyzed asymmetric conjugate addition of alkylzirconium reagents. *Angew. Chem. Int. Edn* **52,** 7995–7999 (2013).

47. Shintani, R., Duan, W.-L. & Hayashi, T. Rhodium-catalyzed asymmetric construction of quaternary stereocenters: ligand-dependent regiocontrol in the 1,4-addition to substituted maleimides. *J. Am. Chem. Soc.* **128,** 5628–5629 (2006).

48. Shintani, R., Tsutsumi, Y., Nagaosa, M., Nishimura, T. & Hayashi, T. Sodium tetraarylborates as effective nucleophiles in rhodium/diene-catalyzed 1,4-addition to β,β-disubstituted α,β-unsaturated ketones: catalytic asymmetric construction of quaternary carbon stereocenters. *J. Am. Chem. Soc.* **131,** 13588–13589 (2009).

49. Hawner, C. et al. Rhodium-catalyzed asymmetric 1,4-addition of aryl alanes to trisubstituted enones: Binap as an effective ligand in the formation of quaternary stereocenters. *Angew. Chem. Int. Edn* **49,** 7769–7772 (2010).

50. Holder, J. C. et al. Mechanism and enantioselectivity in palladium-catalyzed conjugate addition of arylboronic acids to β-substituted cyclic enones: insights from computation and experiment. *J. Am. Chem. Soc.* **135,** 14996–15007 (2013).

51. Mendoza, A., Ishihara, Y. & Baran, P. S. Scalable enantioselective total synthesis of taxanes. *Nature Chem.* **4,** 21–25 (2012).

52. Brown, M. K. & Hoveyda, A. H. Enantioselective total synthesis of clavirolide C. Applications of Cu-catalyzed asymmetric conjugate additions and Ru-catalyzed ring-closing metathesis. *J. Am. Chem. Soc.* **130,** 12904–12906 (2008).

53. Wu, F., Li, H., Hong, R. & Deng, L. Construction of quaternary stereocenters by efficient and practical conjugate additions to α,β-unsaturated ketones with a chiral organic catalyst. *Angew. Chem. Int. Edn* **45,** 947–950 (2006).

54. Poulsen, T. B., Bernardi, L., Alemán, J., Overgaard, J. & Jørgensen, K. A. Organocatalytic asymmetric direct α-alkynylation of cyclic β-ketoesters. *J. Am. Chem. Soc.* **129,** 441–449 (2007).

55. Penon, O. et al. Quaternary stereogenic carbon atoms in complex molecules by an asymmetric, organocatalytic, triple-cascade reaction. *Chemistry* **14,** 4788–4791 (2008).

56. Mase, N., Thayumanavan, R., Tanaka, F. & Barbas, C. F. III Direct asymmetric organocatalytic Michael reactions of α,α-disubstituted aldehydes with β-nitrostyrenes for the synthesis of quaternary carbon-containing products. *Org. Lett.* **6,** 2527–2530 (2004).

57. Kerr, M. S. & Rovis, T. Enantioselective synthesis of quaternary stereocenters via a catalytic asymmetric Stetter reaction. *J. Am. Chem. Soc.* **126,** 8876–8877 (2004).

58. Gao, L., Kang, B. C. & Ryu, D. H. Catalytic asymmetric insertion of diazoesters into aryl-CHO bonds: highly enantioselective construction of chiral all-carbon quaternary centers. *J. Am. Chem. Soc.* **135,** 14556–14559 (2013).

59. Doyle, A. G. & Jacobsen, E. N. Enantioselective alkylation of acyclic α,α-disubstituted tributyltin enolates catalyzed by a {Cr(salen)} complex. *Angew. Chem. Int. Edn* **46,** 3701–3705 (2007).

60. Krautwald, S., Sarlah, D., Schafroth, M. A. & Carreira, E. M. Enantio- and diastereodivergent dual catalysis: α-allylation of branched aldehydes. *Science* **340,** 1065–1068 (2013).

61. Gao, F., Carr, J. L. & Hoveyda, A. H. Copper-catalyzed enantioselective allylic substitution with readily accessible carbonyl- and acetal-containing vinylboron reagents. *Angew. Chem. Int. Edn* **51,** 6613–6617 (2012).

62. Van Veldhuizen, J. J., Campbell, J. E., Giudici, R. E. & Hoveyda, A. H. A readily available chiral Ag-based *N*-heterocyclic carbene complex for use in efficient and highly enantioselective Ru-catalyzed olefin metathesis and Cu-catalyzed allylic alkylation reactions. *J. Am. Chem. Soc.* **127,** 6877–6882 (2005).

63. Dabrowski, J. A., Gao, F. & Hoveyda, A. H. Enantioselective synthesis of alkyne-substituted quaternary carbon stereogenic centers through NHC-Cu-catalyzed allylic substitution reactions with (*i*-Bu)$_2$(alkynylaluminum reagents). *J. Am. Chem. Soc.* **133,** 4778–4781 (2011).

64. Feng, J., Garza, V. J. & Krische, M. J. Redox-triggered C–C coupling of alcohols and vinyl epoxides: diastereo- and enantioselective formation of all-carbon quaternary centers via *tert*-(hydroxy)-prenylation. *J. Am. Chem. Soc.* **136,** 8911–8914 (2014).

65. Behenna, D. C. & Stoltz, B. M. The enantioselective Tsuji allylation. *J. Am. Chem. Soc.* **126,** 15044–15045 (2004).

66. Trost, B. M. & Xu, J. Regio- and enantioselective Pd-catalyzed allylic alkylation of ketones through allyl enol carbonates. *J. Am. Chem. Soc.* **127,** 2846–2847 (2005).
An excellent synopsis of the development of the asymmetric allylic alkylation reaction.

67. Trost, B. M. & Frederiksen, M. U. Palladium-catalyzed asymmetric allylation of prochiral nucleophiles: synthesis of 3-allyl-3-aryl oxindoles. *Angew. Chem. Int. Edn* **44,** 308–310 (2005).

68. Trost, B. M., Xie, J. & Sieber, J. D. The palladium catalyzed asymmetric addition of oxindoles and allenes: an atom-economical versatile method for the construction of chiral indole alkaloids. *J. Am. Chem. Soc.* **133,** 20611–20622 (2011).

69. Trost, B. M., Malhotra, S. & Chan, W. H. Exercising regiocontrol in palladium-catalyzed asymmetric prenylations and geranylation: unifying strategy toward flustramines A and B. *J. Am. Chem. Soc.* **133,** 7328–7331 (2011).

70. Enquist, J. A., Jr & Stoltz, B. M. The total synthesis of (–)-cyanthiwigin F by means of double catalytic enantioselective alkylation. *Nature* **453,** 1228–1231 (2008).

71. Du, C., Li, L., Li, Y. & Xie, Z. Construction of two vicinal quaternary carbons by asymmetric allylic alkylation: total synthesis of hyperolactone C and (–)-biyouyanagin A. *Angew. Chem. Int. Edn* **48,** 7853–7856 (2009).

72. Trost, B. M., Miller, J. R. & Hoffman, C. M. Jr A highly enantio- and diastereoselective molybdenum-catalyzed asymmetric allylic alkylation of cyanoesters. *J. Am. Chem. Soc.* **133,** 8165–8167 (2011).

73. Liu, W.-B., Reeves, C. M., Virgil, S. C. & Stoltz, B. M. Construction of vicinal tertiary and all-carbon quaternary stereocenters via Ir-catalyzed regio-, diastereo-, and enantioselective allylic alkylation and applications in sequential Pd catalysis. *J. Am. Chem. Soc.* **135,** 10626–10629 (2013).

74. Trost, B. M. & Quancard, J. Palladium-catalyzed enantioselective C-3 allylation of 3-substituted-1H-indoles using trialkylboranes. *J. Am. Chem. Soc.* **128,** 6314–6315 (2006).

75. Hills, I. D. & Fu, G. C. Catalytic enantioselective synthesis of oxindoles and benzofuranones that bear a quaternary stereocenter. *Angew. Chem. Int. Edn* **42,** 3921–3924 (2003).

76. Shaw, S. A. et al. Enantioselective TADMAP-catalyzed carboxyl migration reactions for the synthesis of stereogenic quaternary carbon. *J. Am. Chem. Soc.* **128,** 925–934 (2006).

77. Mermerian, A. H. & Fu, G. C. Nucleophile-catalyzed asymmetric acylations of silyl ketene imines: application to the enantioselective synthesis of verapamil. *Angew. Chem. Int. Edn* **44,** 949–952 (2005).

78. DeLorbe, J. E., Jabri, S. Y., Mennen, S. M., Overman, L. E. & Zhang, F.-L. Enantioselective total synthesis of (+)-gliocladine C: convergent construction of cyclotryptamine-fused polyoxopiperazines and a general approach for preparing epidithiodioxopiperazines from trioxopiperazine precursors. *J. Am. Chem. Soc.* **133,** 6549–6552 (2011).

79. Austin, J. F., Kim, S.-G., Sinz, C. J., Xiao, W.-J. & MacMillan, D. W. C. Enantioselective organocatalytic construction of pyrroloindolines by a cascade addition–cyclization strategy: synthesis of (–) flustramine B. *Proc. Natl Acad. Sci. USA* **101,** 5482–5487 (2004).

80. Zhu, S. & MacMillan, D. W. C. Enantioselective copper-catalyzed construction of aryl pyrroloindolines via an arylation–cyclization cascade. *J. Am. Chem. Soc.* **134,** 10815–10818 (2012).

81. Xie, X., Chen, Y. & Ma, D. Enantioselective arylation of 2-methylacetoacetates catalyzed by CuI/*trans*-4-hydroxy-*L*-proline at low reaction temperatures. *J. Am. Chem. Soc.* **135,** 16050–16051 (2008).

82. García-Fortanet, J. & Buchwald, S. L. Asymmetric palladium-catalyzed intramolecular α-arylation of aldehydes. *Angew. Chem. Int. Edn* **47,** 8108–8111 (2008).

83. Taylor, A. M., Altman, R. A. & Buchwald, S. L. Palladium-catalyzed enantioselective α-arylation and α-vinylation of oxindoles facilitated by an axially chiral P-stereogenic ligand. *J. Am. Chem. Soc.* **131,** 9900–9901 (2009).

84. Mitsunuma, H., Shibasaki, M., Kanai, M. & Matsunaga, S. Catalytic asymmetric total synthesis of chimonanthine, folicanthine, and calycanthine through double Michael reaction of bisoxindole. *Angew. Chem. Int. Edn* **51,** 5217–5221 (2012).

85. Zhang, H., Hong, L., Kang, H. & Wang, R. Construction of vicinal all-carbon quaternary stereocenters by catalytic asymmetric alkylation reaction of 3-bromooxindoles with 3-substituted indoles: total synthesis of (+)-perophoramidine. *J. Am. Chem. Soc.* **135,** 14098–14101 (2013).

86. Malcolmson, S. J., Meek, S. J., Sattely, E. S., Schrock, R. R. & Hoveyda, A. H. Highly efficient molybdenum-based catalysts for enantioselective alkene metathesis. *Nature* **456,** 933–937 (2008).
    An excellent example of using desymmetrization to install a key quaternary stereocentre in a structurally complex natural product.
87. Giudici, R. E. & Hoveyda, A. H. Directed catalytic asymmetric olefin metathesis. Selectivity control by enoate and ynoate groups in Ru-catalyzed asymmetric ring-opening/cross-metathesis. *J. Am. Chem. Soc.* **129,** 3824–3825 (2007).
88. Kleinbeck, F. & Toste, F. D. Gold(I)-catalyzed enantioselective ring expansion of allenylcyclopropanols. *J. Am. Chem. Soc.* **131,** 9178–9179 (2009).
89. Phan, D. H. T., Kou, K. G. M. & Dong, V. M. Enantioselective desymmetrization of cyclopropenes by hydroacylation. *J. Am. Chem. Soc.* **132,** 16354–16355 (2010).
90. Souillart, L., Parker, E. & Cramer, N. Highly enantioselective rhodium(I)-catalyzed activation of enantiotopic cyclobutanone C–C bonds. *Angew. Chem. Int. Edn* **53,** 3001–3005 (2014).
91. Shi, B.-F., Zhang, Y.-H., Lam, J. K., Wang, D.-H. & Yu, J.-Q. Pd(II)-catalyzed enantioselective C-H olefination of diphenylacetic acids. *J. Am. Chem. Soc.* **132,** 460–461 (2010).
92. Aikawa, K., Okamoto, T. & Mikami, K. Copper(I)-catalyzed asymmetric desymmetrization: synthesis of five-membered-ring compounds containing all-carbon quaternary stereocenters. *J. Am. Chem. Soc.* **134,** 10329–10332 (2012).
93. Zhang, E., Fan, C.-A., Tu, Y.-Q., Zhang, F.-M. & Song, Y.-L. Organocatalytic asymmetric vinylogous α-ketol rearrangement: enantioselective construction of chiral all-carbon quaternary stereocenters in spirocyclic diketones via

94. semipinacol-type 1,2-carbon migration. *J. Am. Chem. Soc.* **131,** 14626–14627 (2009).
94. Wadamoto, M., Phillips, E. M., Reynolds, T. E. & Scheidt, K. A. Enantioselective synthesis of α,α-disubstituted cyclopentenes by an *N*-heterocyclic carbene-catalyzed desymmetrization of 1,3-diketones. *J. Am. Chem. Soc.* **129,** 10098–10099 (2007).
95. Das, J. P. & Marek, I. Enantioselective synthesis of all-carbon quaternary stereogenic centers in acyclic systems. *Chem. Commun.* **47,** 4593–4623 (2011).
96. Xiao, K.-J. *et al.* Palladium(II)-catalyzed enantioselective C(sp^3)–H activation using a chiral hydroxamic acid ligand. *J. Am. Chem. Soc.* **136,** 8138–8142 (2014).
97. Notte, G. T. New chemical entities entering phase III trials in 2012. *Annu. Rep. Med. Chem.* **48,** 451–469 (2013).
98. Lovering, F. Escape from flatland 2: complexity and promiscuity. *MedChemComm.* **4,** 515–519 (2013).

Author Contributions Both K.W.Q. and L.E.O. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to L.E.O. (leoverma@uci.edu).

ARTICLE

Divergent reprogramming routes lead to alternative stem–cell states

Peter D. Tonge[1], Andrew J. Corso[1,2], Claudio Monetti[1], Samer M. I. Hussein[1], Mira C. Puri[1,3], Iacovos P. Michael[1,4], Mira Li[1], Dong–Sung Lee[5,6,7], Jessica C. Mar[8], Nicole Cloonan[9]†, David L. Wood[9], Maely E. Gauthier[9], Othmar Korn[10], Jennifer L. Clancy[11], Thomas Preiss[11,12], Sean M. Grimmond[9], Jong–Yeon Shin[5,13], Jeong–Sun Seo[5,6,7,13], Christine A. Wells[10], Ian M. Rogers[1,14,15] & Andras Nagy[1,2,15]

Pluripotency is defined by the ability of a cell to differentiate to the derivatives of all the three embryonic germ layers: ectoderm, mesoderm and endoderm. Pluripotent cells can be captured via the archetypal derivation of embryonic stem cells or via somatic cell reprogramming. Somatic cells are induced to acquire a pluripotent stem cell (iPSC) state through the forced expression of key transcription factors, and in the mouse these cells can fulfil the strictest of all developmental assays for pluripotent cells by generating completely iPSC–derived embryos and mice. However, it is not known whether there are additional classes of pluripotent cells, or what the spectrum of reprogrammed phenotypes encompasses. Here we explore alternative outcomes of somatic reprogramming by fully characterizing reprogrammed cells independent of preconceived definitions of iPSC states. We demonstrate that by maintaining elevated reprogramming factor expression levels, mouse embryonic fibroblasts go through unique epigenetic modifications to arrive at a stable, Nanog–positive, alternative pluripotent state. In doing so, we prove that the pluripotent spectrum can encompass multiple, unique cell states.

Somatic cells that have lost their pluripotent properties through the acquisition of differentiation-associated epigenetic marks can be driven to acquire an induced pluripotent cell (iPSC) state by the forced expression of key transcription factors[1]. iPSCs can fulfil the strictest of murine developmental assays, tetraploid embryo complementation[2], to form to all the cells of the embryo proper and the resulting adult animal[3]. During the reprogramming of somatic cells, it is visibly apparent that there exists a spectrum of distinct cell types. The embryonic stem cell (ESC)-like iPSCs capable of generating healthy mice represent just one end of this spectrum. Many studies describe the successful derivation of iPSCs, however, relatively few studies address the fate of cells that do not reprogram to an ESC-like state. It has been reported that somatic cells expressing the four reprogramming factors[1] can stabilize at a Nanog-negative cell state that morphologically resembles ESCs, yet failed to fully acquire an ESC-like expression profile[4–6]. 'Partially reprogrammed cell' has become a term to describe any cell that fails to reprogram to an ESC-like state. However, it is likely that a range of cell types exist, whose stable phenotypes and associated epigenetic profiles are different from ESCs.

For somatic cells to acquire an ESC-like state they require extensive genome-wide remodelling, with epigenetic mechanisms regulating cell state transitions throughout the entire reprogramming process. Incomplete remodelling of the somatic epigenome is associated with transgene-dependent cells[5] and a functional memory of somatic cell origin[7,8]. The modulation of epigenetic regulators such as DNA dioxygenases[9], histone deacetylases[10], H3K36 demethylase (Jhdm1b)[11], H3K27 demethylase (Utx)[12] and H3K9 demethylases[6] greatly influences the efficiency and kinetics of reprogramming towards a ESC-like iPSC state. In particular, vitamin C has been reported to facilitate the transition of cells from a 'partially reprogrammed state' to an ESC-like state[6,13]. In addition to chromatin remodelling, the expression level of reprogramming transcription factors directs cell state. A narrow window of Oct4 expression is required to maintain the ESC state, whereby a twofold perturbation of expression induces cells to transition to a non-ESC state[14]. During reprograming there are two potential sources of Oct4: the transgene, whose expression has to be high at the beginning, and the endogenous gene, which is reactivated during the process of reprogramming. Towards the end of reprogramming the total expression of these two Oct4 sources has to stabilize within the narrow window required by the ESC-like state. Elevated expression of the four reprogramming factors has the potential to direct cell identity to a non-ESC-like state. In agreement, significant changes in global gene expression are observed when the reprogramming factors are shut down[15–17].

Somatic-cell-derived epigenetic marks and the conceivable permutations of reprogramming factor expression levels present a unique opportunity to generate novel cell types. Thus, in an experimental approach unbiased by pre-conceptions of what constitutes a reprogrammed cell we characterize the diversity of cell states that arise during somatic cell reprogramming. We define a Nanog-positive cell state (F-class cells) that is stable, occurs frequently, is dependent on high reprogramming factor expression, in which cells do not form typical ESC-like colonies, exhibits advantageous cell culture properties, and yet demonstrates pluripotency.

[1]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [2]Institute of Medical Science, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [3]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [4]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [5]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul 110-799, South Korea. [6]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 110-799, South Korea. [7]Department of Biochemistry, Seoul National University College of Medicine, Seoul 110-799, South Korea. [8]Department of Systems & Computational Biology, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York 10461, USA. [9]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia. [10]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Queensland 4072, Australia. [11]Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Acton (Canberra), Australian Capital Territory 2601, Australia. [12]Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), New South Wales 2010, Australia. [13]Life Science Institute, Macrogen Inc., Seoul 153-781, South Korea. [14]Department of Physiology, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [15]Department of Obstetrics and Gynaecology, University of Toronto, Toronto, Ontario M5T 3H7, Canada. †Present address: QIMR Berghofer Medical Research Institute, Genomic Biology Lab, 300 Herston Road, Brisbane, Queensland 4006, Australia.

Reprogramming diversity

To extensively characterize the diversity of cell states arising from embryonic fibroblasts, we initiated reprogramming with the doxycycline-inducible *piggyBac* transposon system[18]. Colonies of proliferative cells were picked in a randomized manner, impartial of gene expression and morphological appearance, establishing clonally-derived cell lines (Fig. 1a). Notably, the transgene-expressing cell lines segregated into two distinct cohorts (Fig. 1b), which we had initially classified by morphological appearance as compact colony forming cells (C-class) and fuzzy colony forming cells (F-class). For all 28 cell lines established, the reprogramming genes *Oct4* (also known as *Pou5f1*), *Sox2, Klf4* and *c-Myc* were expressed many fold above ESC levels (Extended Data Fig. 1a), with each clonal cell line exhibiting substantial global gene expression differences when compared to ESCs (Fig. 1b). The majority of genes (67%) that were expressed above ESC levels were also expressed above (>twofold) parental fibroblast levels (Extended Data Fig. 1b, c and

Figure 1 | Fibroblasts reprogram to multiple states. a, Fibroblasts were transfected with Yamanaka factors in four separate *piggyBac* transposons (*pB*) and clonal lines were derived. **b**, Unsupervised hierarchical clustering and sample distance matrix (Pearson correlation) of gene expression at day 16. Phase contrast images representative of F-class (clone 1) and C-class (clone 23) iPS cell lines. Scale bars, 200 μm.

Supplementary Information 1), suggesting that these genes were induced upon reprogramming rather than representing a fibroblast memory. 2,959 differentially expressed genes ($P < 0.01$; false discovery rate (FDR) < 0.05) separated F-class and C-class cells, (Extended Data Fig. 1d, Supplementary Information 2) with the F-class cell lines being particularly intriguing as they expressed Nanog and endogenous Oct4 at ESC levels (Extended Data Fig. 1e, 2a, b), yet did not possess an ESC-like morphology (Fig. 1b). The fuzzy appearance of F-class colonies and low intercellular adhesion was reminiscent of E-cadherin-null ESCs[19–21] and could be attributed to diminished E-cadherin expression (Extended Data Fig. 1e). When mapped to the previously established PluriNet[22] (Extended Data Fig. 1f), F-class cells exhibited significantly reduced expression of many PluriNet genes (*Dnmt3b, Zfp42* and *Tdgf1*), yet they expressed many genes at ESC levels such as *Sall4*, endogenous *Oct4* and *Nanog* (Supplementary Information 3). In addition, the F-class cells expressed transcription factors associated with lineage commitment including the homeobox protein En2, the helix–loop–helix factor Ngn3 and homeobox protein Nkx2.3.

We compared the F-class cells to another well-characterized pluripotent stem cell population, epiblast stem cells (EpiSCs), and found that the F-class cells are transcriptionally distinct (Extended Data Fig. 2c, d). Furthermore, F-class cells could not be generated or maintained in EpiSC media (Extended Data Fig. 2e).

An alternative stem–cell state

Differentially expressed genes ($P < 0.01$; FDR < 0.05) between ESCs and F-class cells are enriched with genes involved in cell adhesion and the extracellular matrix (Fig. 2a, b), which probably contributes to the morphological appearance of F-class cells. Forced expression of Cdh1 induced some cells to acquire an ESC-like morphology; however, it was insufficient for most cells in culture (Extended Data Fig. 3a, b), suggesting that Cdh1 was not the only factor required. Furthermore, elevated Cdh1 expression did not induce the expression of *Esrrb* and *Dppa5*, genes that are downregulated in *Cdh1*-null ESCs[20] (Extended Data Fig. 3a). The F-class gene expression profile remained unchanged upon prolonged culture, with cells maintaining a stable transcriptome and no convergence towards an ESC-like state (Fig. 2c). Independent sub-lines exhibited low variance in gene expression, further demonstrating the stable self-renewal of the F-class cell state (Extended Data Fig. 3c). The absence of interspersed Dppa4-expressing cells suggested that cells do not spontaneously progress to an ESC-like state at a detectable rate (Extended Data Fig. 3d). F-class cells possessed a normal karyotype (Extended Data Fig. 3e) and could be expanded exponentially beyond 40 passages. The cells remained in a transgene-dependent state (Extended Data Fig. 3f), whereby turning off transgene expression induced population-wide differentiation within 48 h, demonstrating that cells had not transformed. The self-renewal of F-class cells was independent of LIF or JAK signalling (Extended Data Fig. 4a, b); furthermore, F-class cells can be generated in media supplemented with JAK inhibitor (Extended Data Fig. 4c–f). F-class cells rapidly proliferated to the extent that, when mixed with ESCs, an initial 1% F-class cells became the dominant cell type (>50%) within three passages (Extended Data Fig. 4g). Stable gene expression, rapid proliferation (Extended Data Fig. 4h) and low intercellular adhesion (Extended Data Fig. 4i) confer F-class cells with highly desirable properties for stirred suspension culture.

Teratomas initiated by pluripotent cells (ESC, ESC-like iPSC and F-class cells) contained well-differentiated (non-dividing) and less differentiated dividing compartments. The teratomas from the F-class cells were indistinguishable from those derived from ESCs, each consisting of complex differentiated tissues representing all three germ layers (Fig. 2d). *In vitro*, removal of doxycycline in serum-free media initiated efficient neural differentiation of F-class cells, generating multiple neuronal subtypes (Extended Data Fig. 5a–c). Differentiation in serum-based media generated cells representative of the mesoderm (α-SMA⁺) and endoderm (FoxA2⁺) lineages (Extended Data Fig. 5d). We then assessed the embryonic developmental potential of F-class cells and found that they

Figure 2 | The F-class state. a, Differentially expressed genes between ESC-like state $n = 4$ and F-class state $n = 6$ (Two-tailed Welchs t-test $P < 0.01$; FDR < 0.05). **b**, Gene ontology term analysis of differentially expressed genes. **c**, Two-way scatter plot comparisons of global gene expression (Illumina BeadArray), blue lines represent fourfold differential threshold. **d**, Histological analysis of teratomas containing differentiated tissues of all three germ layers. Arrowheads denote ciliated epithelia. Scale bars, 100 μm.

do not contribute to the development of chimaeras, nor do they incorporate into blastocysts after injection into the perivitelline space of eight-cell stage embryos (data not shown). In summary, we describe a novel cell state that is distinct from ESCs yet passes criteria used to functionally identify the pluripotent potential of human ESC and iPSC lines, as by the teratoma-forming assay.

Requirement of transgene expression

To determine the influence of transgene expression levels on the establishment of F-class and ESC-like states, we examined three different reprogramming systems: three-factor (3F), which excludes c-Myc; low-expressing four-factor (4F; Col1a1 transgenic secondary system[23]); and high-expressing four factor, 1B secondary system[18] (Extended Data Fig. 6a, b). High-expressing 4F fibroblasts underwent population-wide proliferation and generated distinct colonies within 5 days, which stabilized at a state morphologically and transcriptionally resembling F-class

cells (Extended Data Fig. 6c–e). In contrast, 3F and low-expressing 4F fibroblasts sporadically (<0.1%) gave rise to colonies from day 10 onwards, stabilizing at a state that morphologically and transcriptionally resembled ESCs (Extended Data Fig. 6c–e). During low-expressing 4F reprogramming, no morphologically overt F-class cells were observed at any time point, nor were F-class identifier genes expressed at elevated levels (Extended Data Fig. 6f). These observations suggest a model whereby low-transgene-expressing cells do not generate an F-class cell state (Extended Data Fig. 6g). We found that high four-factor expression can also reprogram adult tail skin fibroblasts to the F-class state (Extended Data Fig. 7a–c).

During somatic cell reprogramming retroviral transgenes become silenced and it is thought that this helps stabilize a fully reprogrammed ESC-like state[24]. Since F-class cells require maintained transgene expression, we questioned whether a retroviral transgene system could give rise to F-class cells. We initially observed rapidly dividing cells possessing an F-class morphology (days 8–16 post-transduction); however, we did not observe these cells beyond day 30. Retrovirus-delivered transgene expression (green fluorescent protein, GFP) was attenuated during transposon-mediated reprogramming to an F-class state and within established F-class cells (Extended Data Fig. 7d, e). We propose that silencing of the retroviral transgenes is not compatible with the F-class cells' requirement for high transgene expression.

To examine the continued requirement of all four reprogramming factors, F-class cells were generated where three factors are constitutively expressed and the fourth factor is doxycycline-inducible. Doxycycline was removed at day 30 and in all four cases turning off the fourth factor induced a rapid loss of proliferation and a flattening of cell morphology (data not shown). Thus, all four reprogramming factors are needed to maintain the F-class state. The consistent inability to obtain F-class cells with 3F reprogramming indicates that elevated c-Myc expression is necessary. We used the TetO-Myc F-class cells, and found that upon doxycycline removal there was a downregulation of genes involved in growth factor activity and positive regulation of transcription (Extended Data Fig. 8a–d), in accordance with a reduced proliferation. Although cells did not transition to an ESC-like state, a number of ESC-associated genes were upregulated (Extended Data Fig. 8c, Supplementary Information 4), supporting the theory that reprogramming factor expression actively suppresses the final acquisition of an ESC-like state[15].

Cell–state transitions

We questioned whether re-expressing the reprogramming factors at high levels in the ESC-like state would induce a transition to the F-class state. Reprogramming factor expression was re-activated in the iPSC line 1B[18] and cells were transferred to media conditions that are conducive to F-class cells but not ESC-like cells: JAK inhibition in the absence of LIF and feeders (Extended Data Fig. 8e, f). Within 48 h, colonies of cells arose that morphologically resembled F-class cells. These cells maintained expression of some ESC-associated genes (*Lin28* and *Dnmt3B*) yet diminished others such as *Dppa5*, *Dnmt3l* and *Cdh1* (Extended Data Fig. 8g). Notably, cells upregulated genes expressed by F-class cells, suggesting that elevated reprogramming transgene expression can induce an F-class-like state, with the starting cell type (ESCs or MEFs) leaving a signature on the F-class cell state.

Next, we investigated whether established F-class cells can be induced to transition to an ESC-like state. Exposure to the DNA methyltransferase inhibitor 5-aza-deoxycytidine (Aza) was toxic at active concentrations (>0.05 μM), while vitamin C (ascorbic acid) supplementation and 2i media failed to induce an ESC-like morphology (Fig. 3a and Extended Data Fig. 9a). In contrast, inhibition of histone deacetylases (HDAC) induced F-class cells to acquire an ESC-like morphology (Fig. 3a) and transcriptional profile (Fig. 3b, Extended Data Fig. 9a). To determine whether HDAC inhibition (HDACi) selects for a sub-population of cells, we exposed twelve newly established subclones to HDACi and found that they acquired an ESC-like morphology and consistently upregulate ESC-like markers (Extended Data Fig. 9b). Furthermore, when
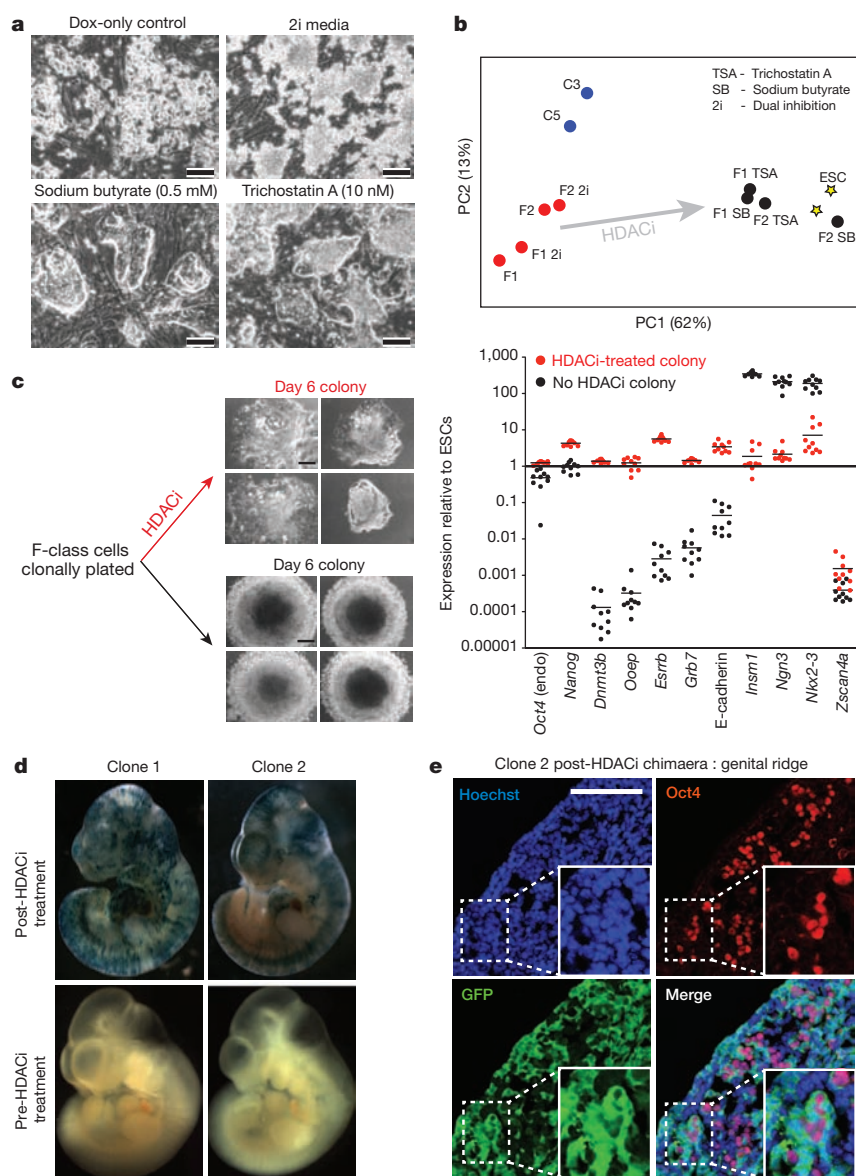
Figure 3 | HDACi induced F-class to ESC-like transition. a, Day 30 F-class cells (clone 1) treated for ten days. Scale bar, 100 μm. **b**, Principal component analysis of 32 cell-state identifier genes (determined by quantitative PCR with reverse transcription, qRT–PCR). C3 and C5 represent C-class clones, F1 and F2 represents F-class clones, maintained in different media. **c**, F-class cells (clone 1) clonally HDACi-treated (10 nM TSA). Scale bar, 250 μm. Each point represents an individual cell colony profiled by qRT–PCR (*n* = 10 biological replicates, 3 technical replicates per colony). *Ngn3* is also known as *Neurog3*. **d**, Chimaeric contribution of HDACi treated F-class cells aggregated with eight-cell stage embryo, visualized by LacZ activity, representative of four embryos. **e**, Genital ridge dissected from chimaeric embryo (*n* = 1). GFP represents HDACi-treated clone2 F-class cells and Oct4 represents the germ cells. Scale bar, 100 μm.

single cells were treated with HDACi, every subsequent colony possessed elevated expression of ESC-associated genes (Fig. 3c). Direct observation of cells by time-lapse microscopy revealed that HDACi treatment decreased cell proliferation (Extended Data Fig. 9c) with no evidence of cell death (Extended Data Fig. 9d). HDACi-mediated acquisition of an ESC-like state was rapid with transcriptionally silent genes upregulated to ESC expression levels within 72 h (Extended Data Fig. 10a–c, Supplementary Information 5). During the first 24 h of HDACi treatment genes with chromatin and cell-division related ontology were upregulated (Extended Data Fig. 10d). The upregulation of chromatin-related factors possibly facilitated the transcriptional activation of further ESC-associated genes. Following HDACi treatment, cells could be maintained as transgene-independent ESC-like cells capable of contributing to chimaeras and the germ line (Fig. 3d, e). This was not possible before HDACi treatment.

Epigenetic forces contribute to F–class state

To identify the epigenetic landmarks associated with the establishment of the F-class cell state, we exploited a high-resolution genome-wide resource that profiles fibroblast reprogramming at the molecular level to both F-class and ESC-like states[25]. Doxycycline-induced high-level reprogramming factor expression directs 1B secondary fibroblast reprogramming

to an F-class transcriptional state (Extended Data Fig. 10e)[25–28]. Comparison of primary F-class cell lines and ESC-like cell lines identified 86 genes that exhibited substantial (>fivefold) differential expression (Fig. 4a). For these genes we assessed the status of three major chromatin marks; the activating histone H3K4 trimethylation (H3K4me3), the suppressing histone H3K27 trimethylation (H3K27me3)[25] and CpG methylation[27] (Supplementary Information 6). Transcriptional activity of 72 of the 86 genes (79%) correlated (Pearson correlation coefficient |*r*| > 0.5) with at least one epigenetic mark (Fig. 4b). The upregulation of F-class state identifiers, such as *Nkx2-3* and *Insm1* (Fig. 4c, d), was associated with an active loss of H3K27me3 during the reprogramming process, fitting the model that the F-class state is not an intermediate reprogrammed state but a distinct cell state achieved through active epigenetic changes. Further substantiating this is the observation that genes associated with the ESC-like state (*Gbx2*, *Lefty1*, *Cldn6*) acquired hypermethylation at their genomic loci (Fig. 4e), which is uncharacteristic of the ESC-like state. We further validated a subset of differentially methylated regions (DMRs) within primary F-class cells (Fig. 4f). In summary, fibroblast reprogramming to the F-class state is governed by multiple epigenetic marks, whereby active epigenetic modifications direct cell identity away from both fibroblast and ESC-like state, and repressive epigenetic marks are inherited from the parental cell type (fibroblasts).
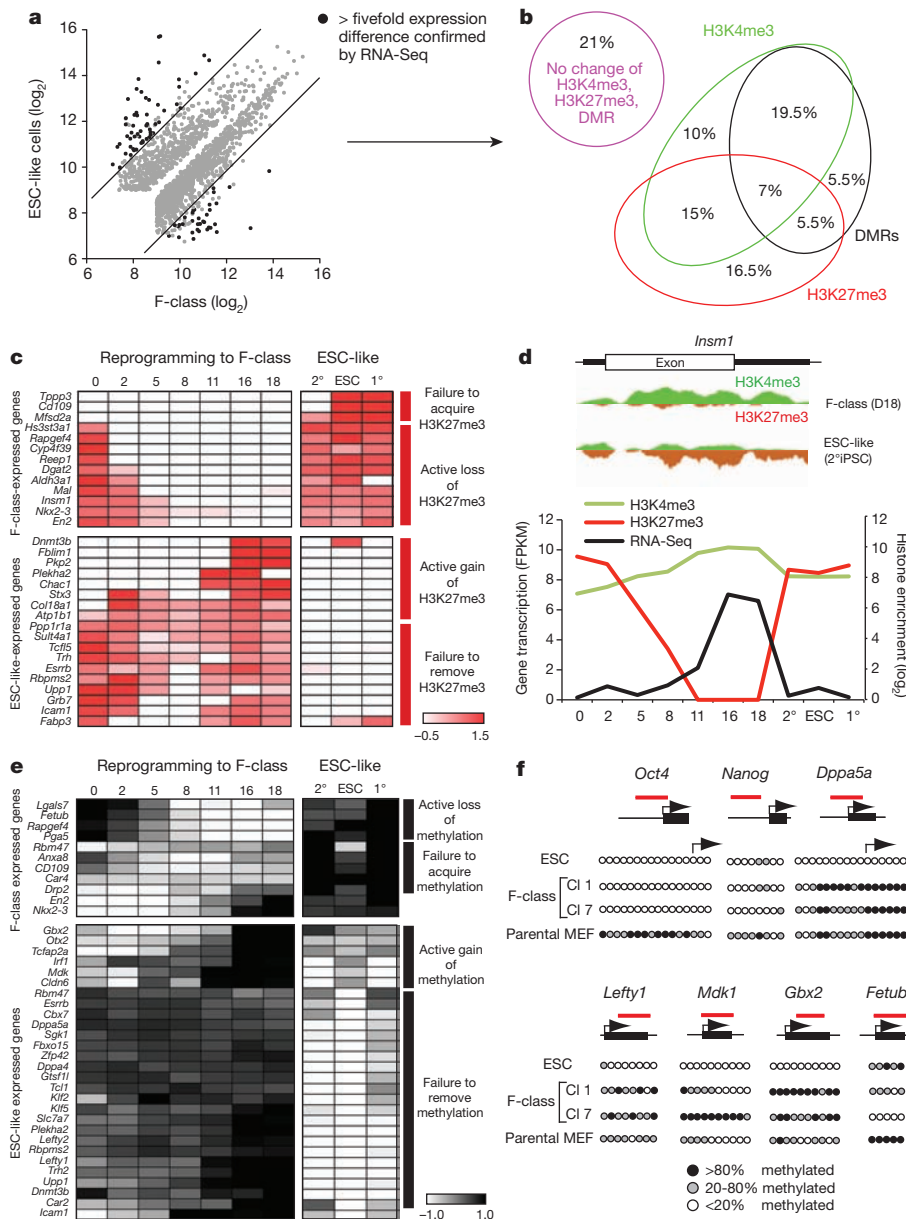
Figure 4 | Epigenetic marks steer reprogramming trajectory. a, Differentially expressed genes in primary lines (Welch *t*-test $P < 0.01$, FDR < 0.01). Black data points depict >fivefold difference in 1B secondary reprogramming system of F-class and ESC-like cells[25]. Lines depict fivefold threshold. **b**, Euler diagram depicting genes (black points in **a**) whose transcriptional activity corresponds with differential epigenetic marks in 1B secondary reprogramming system derived F-class cells[25] ($n = 1$). **c**, Unsupervised clustering heat map of H3K27me3 marks identified in **b**. **d**, Transcription and histone modifications at the genomic locus of F-class identifier *Insm1*. **e**, Unsupervised clustering heat map of differentially methylated regions identified in **b**. **f**, Differentially methylated regions observed in the secondary reprogramming system confirmed in primary F-class cells.

Discussion

In this study, we observed that reprogramming somatic cells, in the presence of elevated reprogramming factor expression, could stabilize at a Nanog-positive fuzzy colony forming (F-class) state. Previous studies may have overlooked this state as the F-class cells highly express Nanog without completing one of the early reprogramming events, the mesenchymal to epithelial transition[16,29]. Chan and colleagues previously described a human reprogrammed cell type (type II cells) that is Nanog-positive and persists in a state that represents an intermediate stage of somatic cell reprogramming[30]. In contrast to the human type II cells, the murine F-class cells do not morphologically resemble ESCs, nor do they transcriptionally or epigenetically represent an intermediate cell state that reprogramming cells transit through as they acquire ESC-like state. Two central observations support the notion that the F-class cell state is not representative of an intermediate state. First, F-class cells upregulate a cohort of genes that were not observed during reprogramming without c-Myc (3F) or with low-level four-factor (Oct4, Klf4, Sox2 and c-Myc) expression. Second, the expression of these genes in F-class cells is associated with the loss of repressive epigenetic marks (H3K27me3 and/or DNA methylation) that are typically present in the parental fibroblasts and the ESC-like state. The loss of these repressive marks

suggests that, during sustained reprogramming factor expression, cell identity is diverted away from the molecular pathway that leads to an ESC-like state (Fig. 5). This is further supported by the observation that ESC-associated genes (*Lefty1*, *Cldn6*, *Gbx2*) actually acquire inhibitory DNA methylation in the F-class state. To our knowledge, this is the first report to identify dynamic epigenetic changes that actively propel reprogramming cells towards an alternative pluripotent cell state.
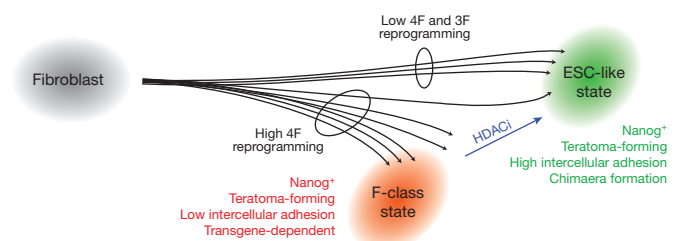
Figure 5 | Schematic representation of cell-state transitions during reprogramming. HDACi denotes histone deacetylation inhibition, 4F denotes the four Yamanaka factors, 3F denotes the four Yamanaka factors minus c-Myc.

In conclusion, the F-class cells represent an acquired state and not an intermediate state that all reprogramming cells transition through on the way to an ESC-like state.

We propose that the F-class cell state is stably maintained as a consequence of high reprogramming factor expression and multiple epigenetic determinants. Through elevated expression of the four reprogramming factors we showed that F-class cells could be generated from both fibroblasts and ESC-like iPSCs. Notably, the cell type of origin leaves distinct signatures on the resultant F-class cells, as an imprint of their respective origin.

The ability to reprogram cells to novel cell states, such as the F-class state, can be harnessed to create a variety of artificial cells that possess desirable properties for regenerative medicine and drug discovery, such as the ability for scalable expansion in bioreactors and reproducible differentiation. ESCs are themselves an artificial *in vitro* cell state, captured during a brief developmental window and require specific culture conditions for their maintenance. The F-class cell state can be considered to be a distant pluripotent relative of the ESC state. The frequency at which F-class cells arise in transposon-based reprogramming, in combination with their advantageous properties, presents the opportunity to study and utilize a novel pluripotent cell type in biology, medical research and future medicine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90,** 8424–8428 (1993).
3. Zhao, X.-Y. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461,** 86–90 (2009).
4. Fussner, E. *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* **30,** 1778–1789 (2011).
5. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
6. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genet.* **45,** 34–42 (2013).
7. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).
8. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnol.* **28,** 848–855 (2010).
9. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14,** 512–522 (2014).
10. Mali, P. *et al.* Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells* **28,** 713–720 (2010).
11. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9,** 575–587 (2011).
12. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488,** 409–413 (2012).
13. Esteban, M. A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6,** 71–79 (2009).
14. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
15. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11,** 769–782 (2012).
16. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
17. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
18. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).
19. Soncin, F. *et al.* Abrogation of E-cadherin-mediated cell–cell contact in mouse embryonic stem cells results in reversible LIF-independent self-renewal. *Stem Cells* **27,** 2069–2080 (2009).
20. Soncin, F. *et al.* E-cadherin acts as a regulator of transcripts associated with a wide range of cellular processes in mouse embryonic stem cells. *PLoS ONE* **6,** e21463 (2011).
21. Larue, L. *et al.* A role for cadherins in tissue formation. *Development* **122,** 3185–3194 (1996).
22. Müller, F.-J. *et al.* Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **455,** 401–405 (2008).
23. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26,** 916–924 (2008).
24. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
25. Hussein, S. M. I. *et al.* Genome-wide characterization of the routes to induced pluripotency. *Nature* http://dx.doi.org/10.1038/nature14046 (2014).
26. Benevento, M. *et al.* Proteome adaptation in cell reprogramming proceeds via distinct transcriptional networks. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6613 (2014).
27. Lee, D. S. *et al.* An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6619 (2014).
28. Clancy, J. L. *et al.* Small RNA changes *en route* to distinct cellular states of induced pluripotency. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6522 (2014).
29. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
30. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnol.* **27,** 1033–1037 (2009).

METHODS

Cell culture. All cell lines were established in-house with full pathogen testing performed and maintained in a mycoplasma-free facility. Mouse embryonic fibroblasts (MEF) were isolated as previously described[18]. 15.5 days post coitum ROSA26-rtTA-IRES-GFP embryos (JAX 005572)[31] were decapitated, eviscerated, dissociated with 0.25% trypsin, 0.1% EDTA and plated in DMEM, 10% FBS, penicillin–streptomycin and GlutaMAX. MEFs were reprogrammed within 4 passages of derivation. Tail-tip fibroblasts (TTFs) were obtained from 8-week-old mice. Tail-tips were mechanically dissociated with 0.25% trypsin and 1,000 U ml^{-1} collagenase (Type XI-S).

A standardised transfection protocol was established to electroporate fibroblasts (Neon, Invitrogen) with *piggyBac* transposons encoding the four reprogramming factors. In brief, 2×10^6 MEFs were electroporated with 4 µg of plasmid (0.5 µg PBase transposon and 3.5 µg factors), using optimized parameters (2 pulses, 1,200 V). Electroporated fibroblasts were plated in serum-based mouse ESC media[32] supplemented with 1.5 µg ml^{-1} doxycycline on gelatinized (0.1%) plates, at a density of 1.5×10^4 cells per cm^2. Cells were fed every three days with doxycycline-containing media (1.5 µg ml^{-1}). Colonies were clonally picked and expanded in a 96-well format. Unless stated otherwise, clonal cell lines were maintained in mouse ESC media supplemented with 1.5 µg ml^{-1} doxycycline. ROSA26-rtTA-IRES-GFP ESCs were used as control cells. 2i media conditions represent serum-free media consisting of DMEM:F12 supplemented with 15% Knockout serum replacement (Gibco), 3 µM CHIR99021 (GSK3β inhibitor) and 1 µM PD0325901 (MEK inhibitor) as previously described[33].

Transgene independent ESC-like iPSCs were obtained from F-class cells by exposure to sodium butyrate (0.25 mM) for seven days (plus doxycycline). Cells were then maintained in 2i media in the absence of sodium butyrate (plus doxycycline) for five days and then doxycycline was removed. Cells were furthermore maintained in either serum-based ESC media or 2i media.

EpiSCs were maintained in X-vivo base media (Lonza) supplemented with 10 mM β-mercaptoethanol (Sigma), 1 mM MEM-NEAA (Invitrogen), 2 mM GlutaMAX (Invitrogen), 20 ng ml^{-1} Activin A (R&D Systems), and 20 ng ml^{-1} basic fibroblast growth factor (R&D Systems). EpiSCs were passaged every 3–4 days as single cells in TrypLE (Invitrogen) and plated on wells pre-coated with Matrigel.

For retrovirus mediated reprogramming, retroviral packaging of pMX constructs and subsequent transduction of cells was performed as previously described[34].

Stirred suspension culture. Adherent cells were trypsinized and seeded into spinner flasks at 2×10^4 cells per ml. 30-ml culture volumes were maintained at constant stirring speed of 85 r.p.m. at 37 °C and 10% CO$_2$. Every three days cell numbers were quantified and suspension cultures reset to 2×10^4 cells per ml. One-half of the culture medium was replaced every two days.

***In vitro* neural differentiation.** Cells were plated on geltrex (1:100 PBS dilution) coated plates at 5,000 cells per cm^2. 24 h after plating cells, ESC media was changed to serum-free media that consisted of DMEM:F12 supplemented with N2 (Gibco), B27 (Gibco), and 4 µg ml^{-1} insulin. Doxycycline was removed by washing cells three times with PBS to remove all traces of doxycycline. Differentiation media was changed every three days.

Diploid aggregation generation of chimaeras. Cells were maintained for two passages in 2i media with cell clumps of ~8–15 cells collected from gelatinized dishes by gentle trypsinization. For diploid chimaeras, 2.5 d.p.c. Hsd:ICR(CD-1) or C57BL/6 embryos were aggregated with *in-vitro*-derived cell clumps and cultured overnight at 37 °C in 5% CO$_2$ in KSOM medium[33]. All embryos were transferred into pseudopregnant recipient ICR females 24 h later. For LacZ detection, pregnant dams were fed doxycycline food and water (0.2 mg ml^{-1} doxycycline; 5% sucrose in water) 24 h before dissection to activate β-geo expression in iPSC-derived cells. All mouse procedures were performed in accordance with Toronto Centre for Phenogenomics animal care committee.

LacZ staining. As described in ref. 18 cells and embryos were fixed with 0.25% glutaraldehyde, rinsed in wash buffer (2 mM MgCl$_2$, 0.01% sodium deoxycholate, and 0.02% Nonidet-P40 in PBS) and stained overnight (~16 h) in LacZ staining solution: 20 mM MgCl$_2$, 5 mM K$_3$Fe(CN)$_6$, 5 mM K$_4$Fe(CN)$_6$ and 1 mg ml^{-1} X-gal in PBS. Embryos were embedded in paraffin, sectioned and counterstained with neutral red.

Teratoma formation. Cells were trypsinized and suspended in DMEM:Matrigel mix (1:1) with 100 µl of 1×10^6 cells injected subcutaneously into the dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J females, 6 weeks of age) anaesthetized with isoflurane. 4–6 weeks after injection, teratomas were dissected and fixed overnight in 4% formalin. Tissue was embedded in paraffin, sectioned and stained with haematoxylin and eosin.

Immunostaining and flow cytometry. Cells were washed once with PBS, fixed in 4% PFA for 15 min at room temperature and permeabilized with 0.1% Triton X-100 in PBS for 10 min. Primary antibody was added overnight at 4 °C: anti-α-SMA (C6198, Sigma), anti-Nanog (RCAB0002P, Reprocell), anti-DPPA4 (AF3730, R&D Systems), anti-FoxA2 (ab40874, Abcam) anti-SSEA1 (MAB4301, Millipore), anti-Sox2 (MAB2018, R&D Systems), anti-Oct3/4 (611203, BD), anti-GFP (6673, Abcam), anti-βIII-tubulin (TUJ1, Covance), anti-tyrosine hydroxylase (AB152, Millipore), anti-VGAT (131103, SYSY), anti-VGLUT1 (135302, SYSY). Secondary antibody (Jackson immune research cy3 IgG, 1:200; Alexa488 IgG or IgM, 1:400; Alexa594 IgG, 1:400) was added for 1 h at room temperature. Cell nuclei were stained with Hoechst 33342 (5 µg ml^{-1}) for 15 min.

Flow cytometry. Cells were trypsinized and fixed in 4% PFA for 15 min at room temperature. Cells were washed and then stained with 0.1% Triton X-100 in PBS (2% FBS), incubated with primary antibody (Nanog 1:200) for 1 h on ice, washed twice in PBS (2% FBS), incubated with secondary antibody for 30 min on ice, washed twice and resuspended in PBS with 2% FBS for analysis on a FACS-Calibur. Cells were gated on the basis of forward scatter and side scatter.

Cell viability assay. Cell samples were trypsinized, resuspended in Annexin V buffer (10 mM HEPES, 140 mM NaCl, and 2.5 mM CaCl$_2$, pH 7.4) and then incubated with Sytox AADvanced for 5 min and Annexin V for 10 min. Cellular fragments and debris were excluded from analysis using forward-scatter and side-scatter selection.

G-band karyotyping. G-banding was performed on actively dividing cells at the TCAG facility (Toronto, Canada). Cells were incubated with 0.2 µg ml^{-1} colcemid for 2 h at 37 °C and dissociated with 0.25% trypsin-EDTA. After pipetting a single-cell suspension was resuspended in pre-warmed (37 °C) 75 mM KCl for 15 min. Cells were then fixed with methanol:glacial acetic acid (1:3) and dropped onto glass slides. The slides containing cells were stained in Giemsa solution for 3 min, with 20 metaphases counted and scored for karyotyping.

Quantitative RT–PCR. Cells for RNA preparation were passaged on gelatin-coated plates. Total RNA was extracted from cells using a RNeasy kit (Qiagen). 1 µg of DNase treated RNA was used as template to generate cDNA by QuantiTect reverse transcription kit (Qiagen). For quantitative RT–PCR we used LuminoCt SYBR Green qPCR ReadyMix (Sigma) with JANUS automated liquid handling robot (Perkin-Elmer) loading the 384-well plates for RT-qPCR. 384 plates were run on a CFX384 (Bio-Rad) with an annealing temperature of 58 °C for all primers. Primer pairs were all assessed for efficiency and melt curves performed. All PCR reactions were performed in triplicate. Primer sequences are listed in Supplementary Information 7.

Illumina BeadChip. Total RNA was assessed for quality and quantity on a Bioanalyzer and global gene expression profiling performed with the Illumina microarray. Purified and labelled RNA was hybridized to MouseRef-8 v2 expression BeadChips (Illumina) according to the manufacturer's instructions. Bead intensities were mapped to gene information using BeadStudio 3.2 (Illumina). Background correction was performed using the Affymetrix Robust Multi-array Analysis and data log$_2$-scaled with gene expression quantile normalized in the lumi package of Bioconductor.

Bisulphite sequencing. Bisulphite conversion was performed on genomic DNA sample (1 µg) using the EpiTect Bisulfite Kit (QIAGEN). Bisulphite-treated genomic DNA was amplified by EpiTaq HS (Takara) using previously published bisulphite-specific primers[35] and novel primers (Supplementary Information 4), with a PCR protocol consisting of an initial 1 min denaturation step followed by 35 cycles of 95 °C for 15 s, 55 °C for 30 s and 72 °C for 30 s. The resultant PCR amplicons were cloned in to pGemTeasy and sequenced at the Centre for Applied Genomics (Toronto, Canada).

Statistical analysis. Unless otherwise stated, all data presented are representative of at least three independent experiments. Hierarchical clustering, principal component analysis and gene distance matrices were performed with Multiexperiment Viewer. Statistical analysis was performed with either Prism (Graphpad) or Multiexperiment viewer (http://www.tm4.org/index.html). Gene ontology term analysis was performed with DAVID (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov). Gene network association analysis was performed with GeneMANIA (http://www.genemania.org). Genes in the network analysis were chosen based on their membership of the PluriNet network[22] and statistically significant differential expression between F-class samples and ESC samples. Differential expression was assessed using the limma package, *P* values were adjusted using the Benjamini–Hochberg method and significance cut-off set at 0.05.

31. Belteki, G. *et al.* Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res.* **33**, e51 (2005).
32. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. R. *Manipulating the Mouse Embryo: A Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2003).
33. Gertsenstein, M. *et al.* Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE* **5**, e11260 (2010).
34. Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).

35. Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6,** 34 (2006).

36. Han, D. W. *et al.* Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell* **143,** 617–627 (2010).
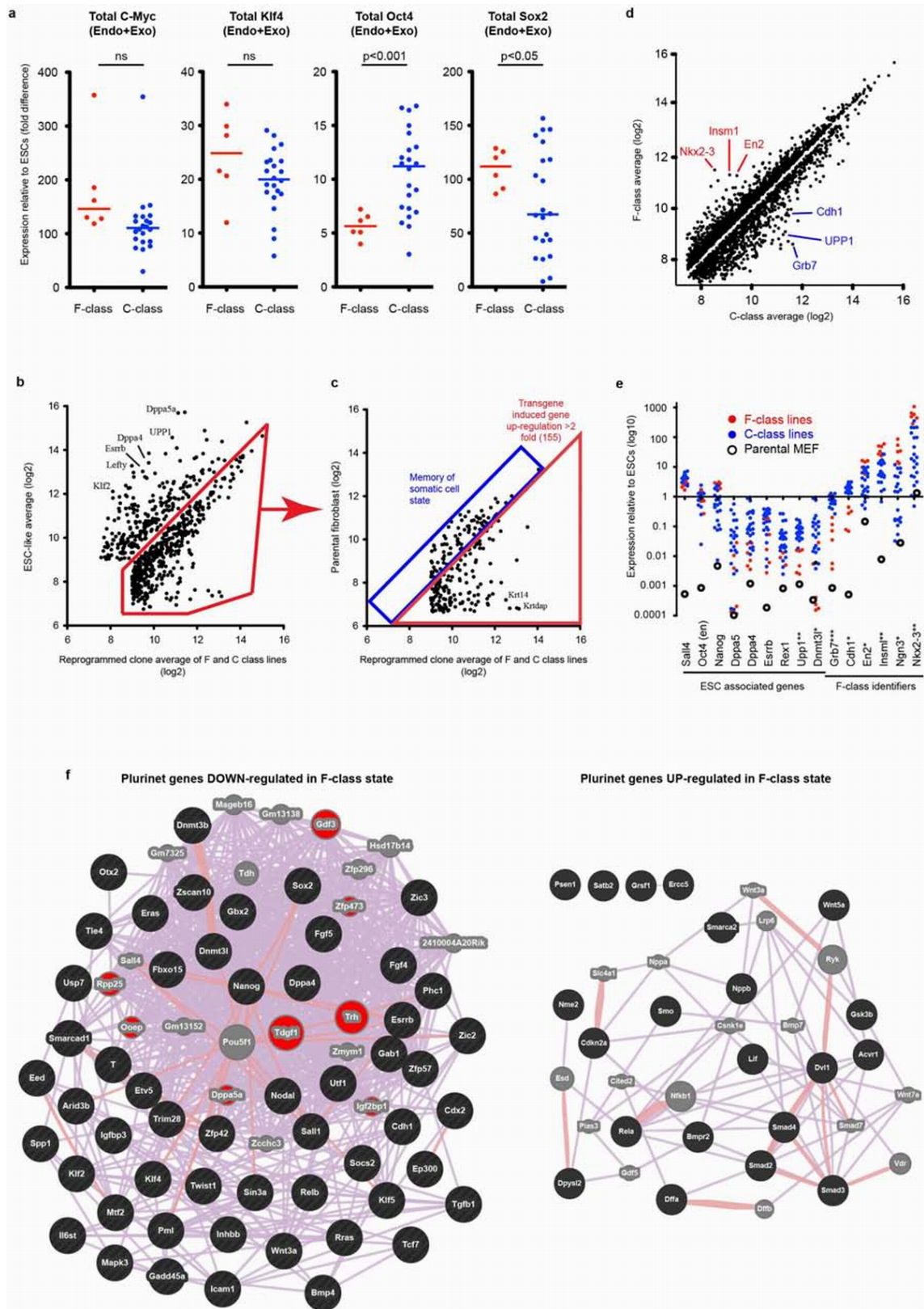
Extended Data Figure 1 | Expression profile of F-class cells. a, Quantitative RT–PCR analysis of total reprogramming factor expression in day 16 F-class ($n = 6$) and C-class ($n = 22$), non-parametric *t*-test. **b**, Differentially expressed genes (two-tailed Welch *t*-test $P < 0.01$, FDR < 0.01) between transgene-expressing reprogrammed lines ($n = 28$) and ESC-like lines ($n = 3$). **c**, Genes highly expressed in **b** compared against parental fibroblasts. Genes $>$twofold higher than fibroblasts classified as reprogramming induced. **d**, Scatter plot of differentially expressed genes (Welch's *t*-test $P < 0.01$; FDR < 0.05).

e, Quantitative RT–PCR profiling of cells in **a**. Non-parametric *t*-test between the F- and C-class lines ($n = 28$) ; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **f**, Expression of PluriNet genes were compared between ESC-like state and F-class state (P values < 0.05, adjusted using the Benjamini–Hochberg method). GeneMANIA interaction network of known gene co-expression and physical interactions. Black nodes represent input genelist, grey nodes represent connecting genes, red nodes represent non-PluriNet genes identified by GeneMANIA that are downregulated in F-class cells.
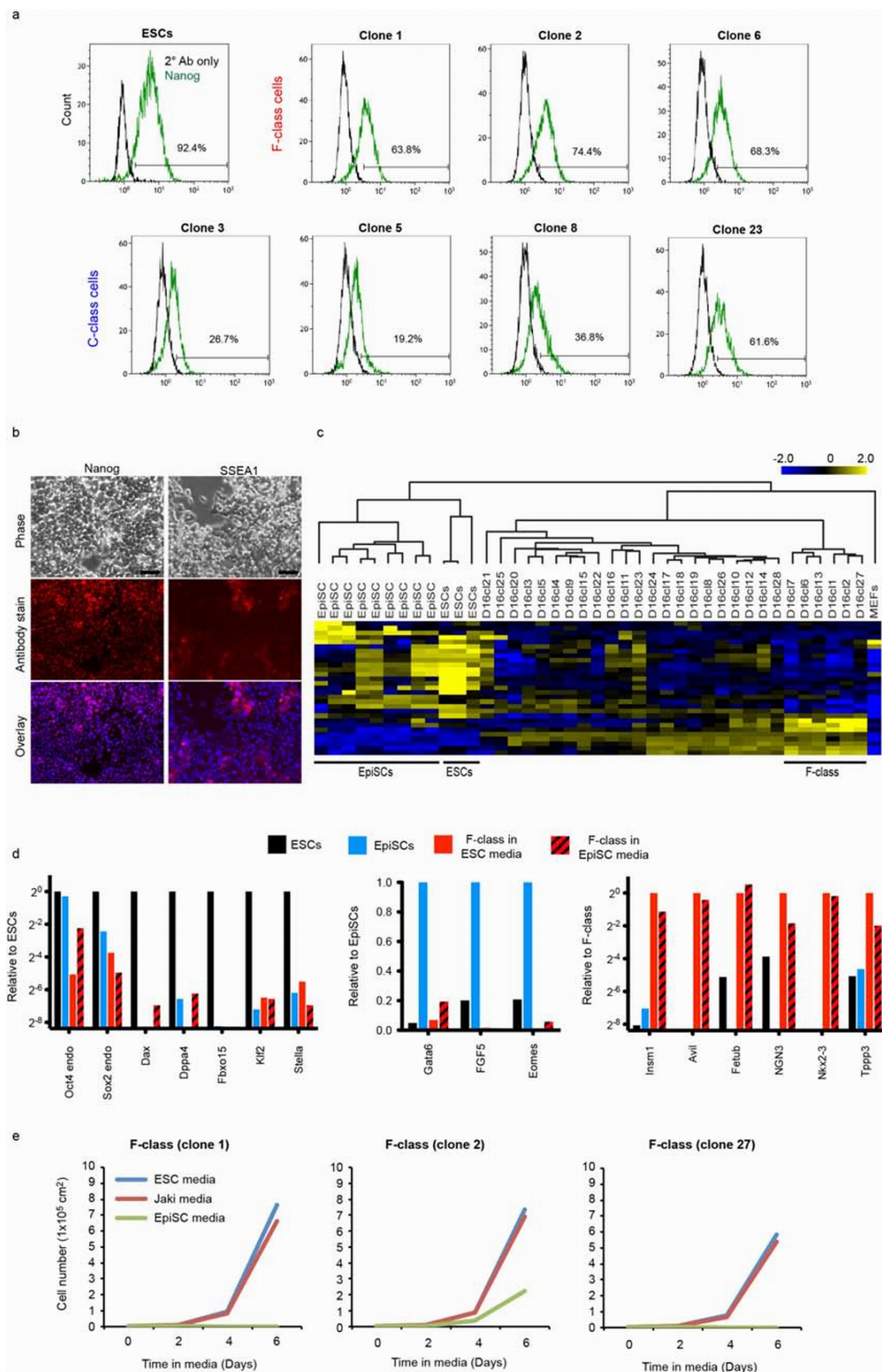
Extended Data Figure 2 | Comparison to epiblast stem cells. a, Flow cytometric analysis of Nanog expression in F- and C-class primary cell lines after 21 days of transgene expression. Graphs show one of $n = 2$ experiments. **b,** Immunofluorescent staining of F-class cells (clone 2) after 30 days of transgene expression. Blue represents Hoechst DNA stain. Scale bars, 100 µm. **c,** Unsupervised hierarchical clustering of gene expression. EpiSC and ESC populations from ref. 36, with all other cell lines described in Fig. 1b. **d,** Quantitative RT–PCR analysis of F-class cells (day 30) grown in EpiSC media for 7 days. Graphs show one of $n = 2$ biological replicates, with 3 technical replicates each. **e,** Proliferation of established F-class cells (day 30) plated in different media compositions, 1,000 cells plated per cm^2, $n = 3$ technical replicates from one experiment.
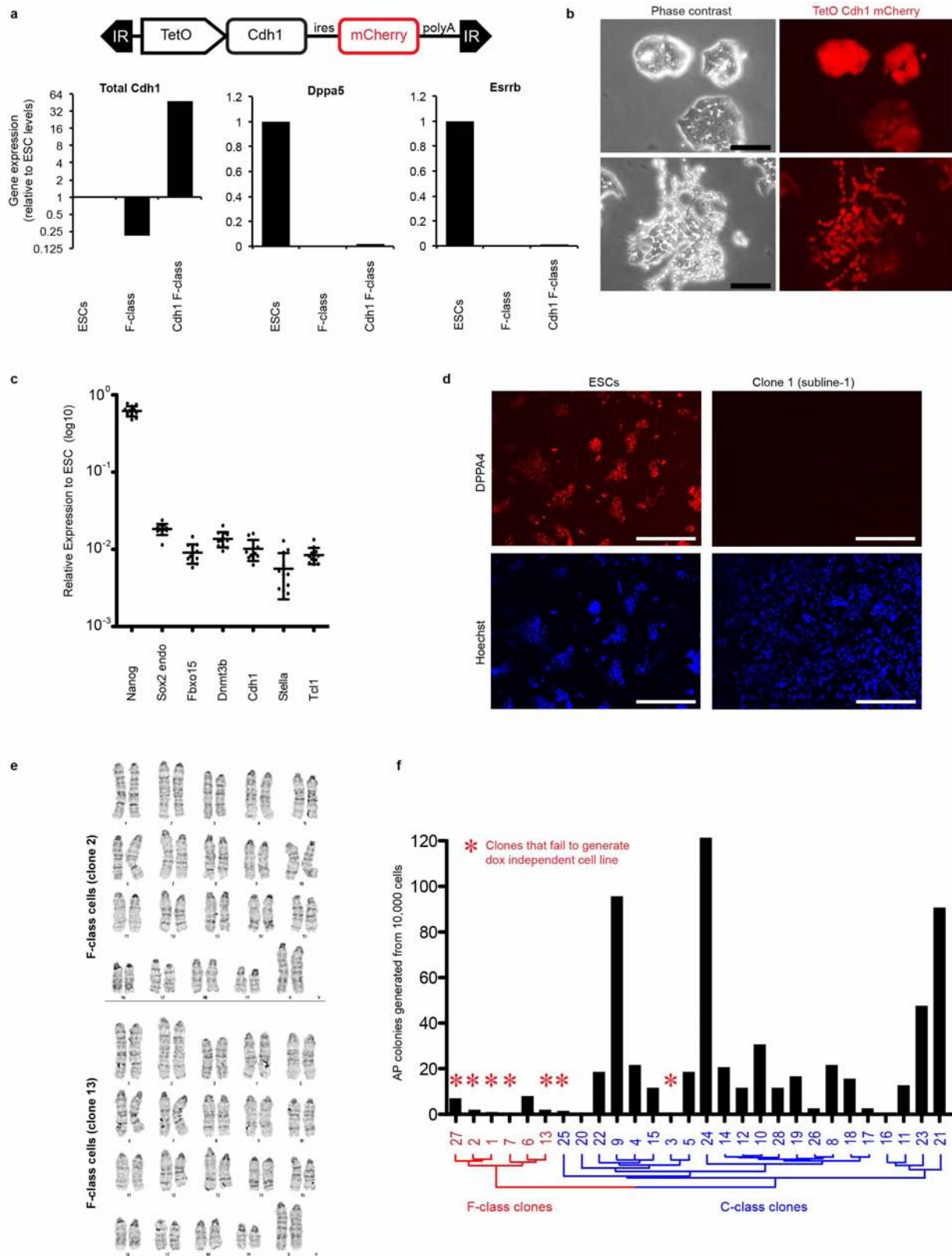
Extended Data Figure 3 | A stable stem-cell state. a, Schematic of Cdh1 overexpressing sleeping beauty transposon. IR depicts sleeping beauty inverted repeats. Quantitative RT–PCR of gene expression after 7 days Cdh1 overexpression. $n = 3$ technical replicates from one experiment. **b**, Images of Cdh1 overexpressing F-class cells. Scale bars, 100 μm. **c**, Quantitative RT–PCR analysis of 12 sub-lines derived from clone 1 F-class cells. Average ± s.d.

d, DPPA4 immunofluorescence of clone 1 (sub-line-1) after 30 days of transgene expression. Scale bars, 200 μm. **e**, G-banded karyotype on diploid metaphases of F-class clones. **f**, Ability to maintain a reprogrammed state in the absence of transgene expression, doxycycline removed after day 21, $n = 3$ technical replicates from one experiment. Clonal lines ordered as in Fig. 1b.
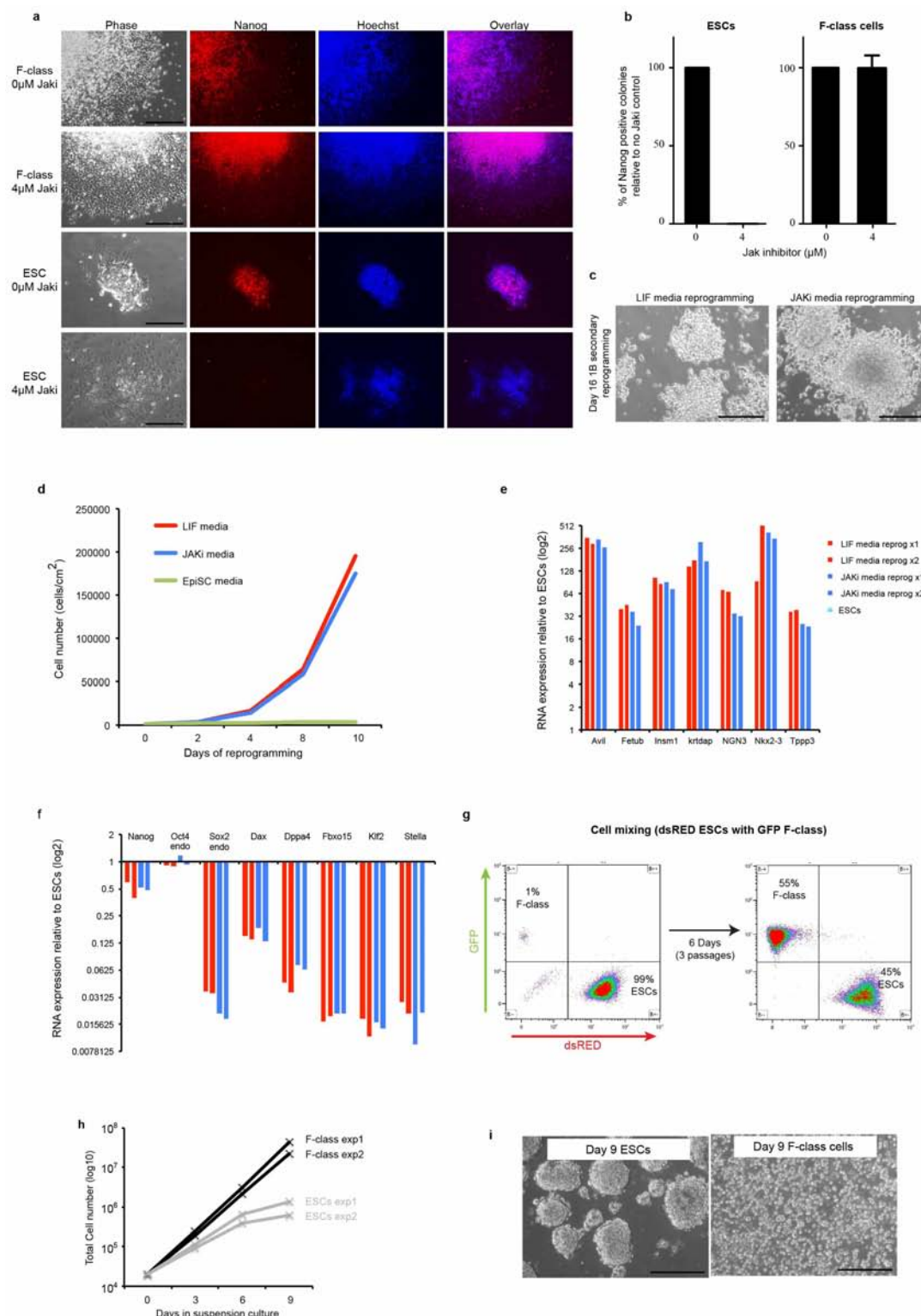
Extended Data Figure 4 | F-class expansion in absence of LIF signalling.
a, Nanog immunofluorescence of single-cell-derived colonies (Day 5). **b**, Clonal efficiency of F-class cells and ESCs treated with JAK inhibitor (data shown is the mean from $n = 3$ biological replicates, with 3 technical replicates each, average ± s.d.). **c**, 1B secondary fibroblast reprogramming[25] initiated by doxycycline treatment of fibroblasts in either JAKi-supplemented media (no LIF) or LIF-supplemented media (standard serum-based ESC media). Scale bars, 200 μm. **d**, Cell expansion of **c** during 10 days of reprogramming (data shown is the mean from 3 technical replicates from one experiment). **e**, **f**, Gene expression analysis (qRT–PCR) of Day 16 reprogramming in JAKi and LIF media (**c**). Assessment of F-class markers (**e**) and ESC markers (**f**) (data shown is the mean from $n = 2$ biological replicates with 3 technical replicates each). **g**, DsRed ESCs were mixed with GFP F-class cells. Flow cytometric analysis of population composition before and after passaging. **h**, Proliferation of F-class and ESC cells grown as suspension culture. **i**, Phase contrast image of cells grown in suspension for 9 days. Scale bars, 200 μm.
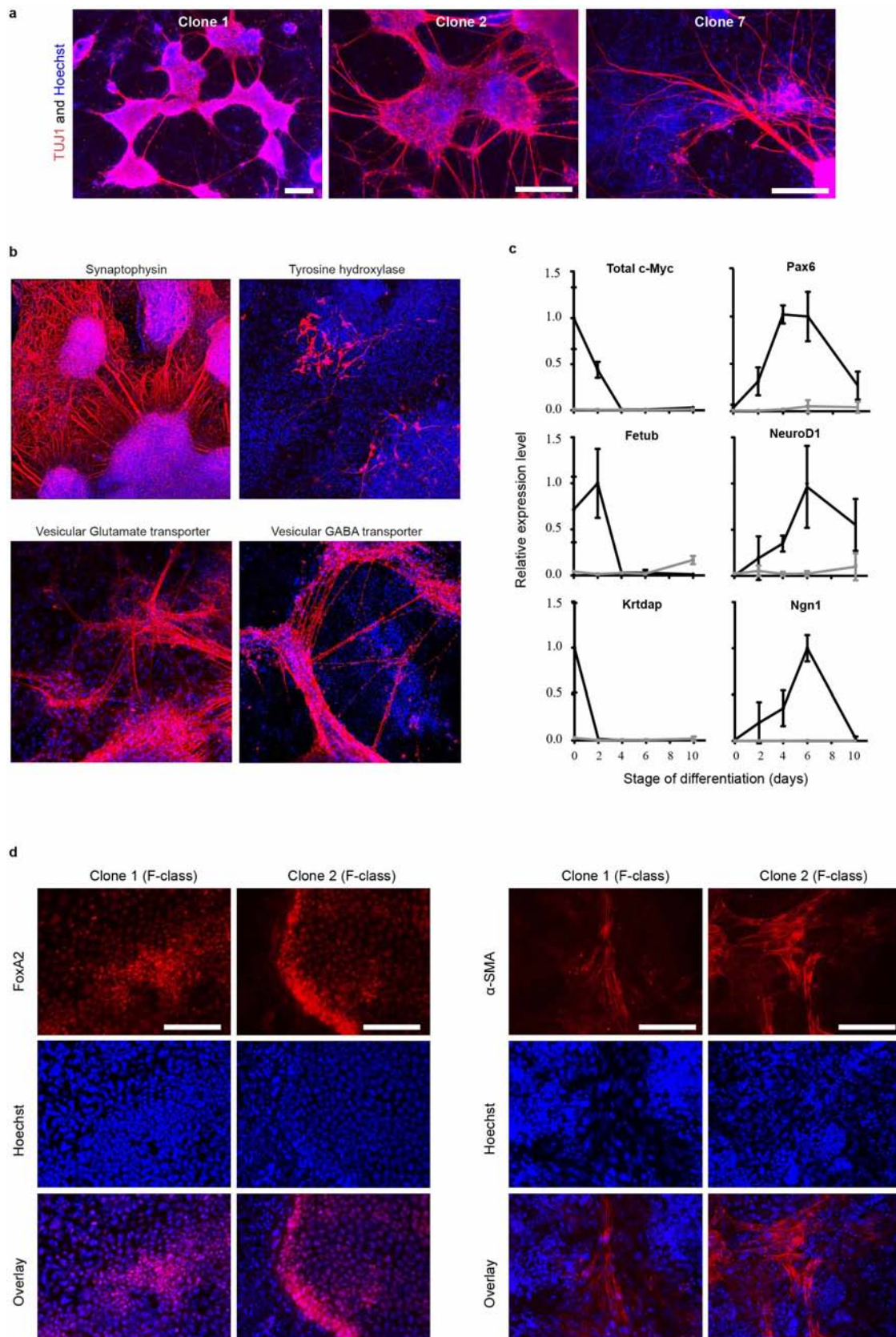
Extended Data Figure 5 | *In vitro* differentiation to three germ layers.
a, TUJ1-positive neurons generated by F-class cells upon doxycycline withdrawal in serum-free media (day 30). **b**, Multiple neuronal subtypes generated by F-class cells, (clone 1, sub-line 1). **c**, Quantitative RT–PCR analysis of gene expression during neural differentiation. Clone 1 F-class cells (black line) in comparison to ESC differentiation (grey line). 3 biological replicates (average ± s.d.). **d**, Doxycycline withdrawal induced differentiation of day 35 F-class cells in 15% serum-based media for 8 days. Immunofluorescent staining of cells representing endoderm (FoxA2) and mesoderm (α-SMA). Scale bars, 200 μm.
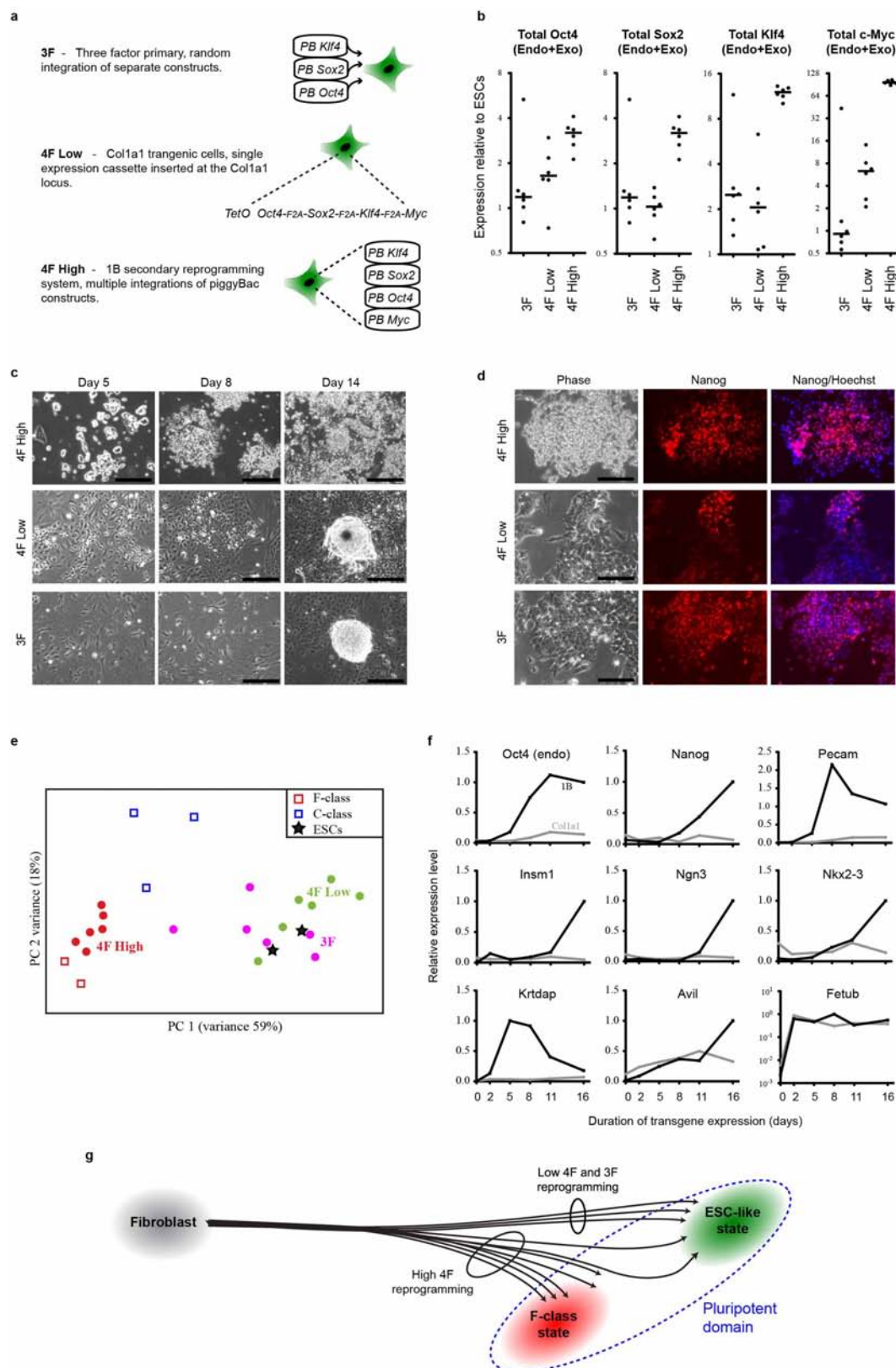
Extended Data Figure 6 | Transgene expression levels direct reprogramming. a, Schematic representation of three assessed reprogramming systems. **b**, Quantitative RT–PCR of factor expression of clonal lines, day 25–day 35. Each point represents a clonal reprogramming colony with $n = 6$ biological replicates and 3 technical replicates each. **c**, Representative images of transgene-expressing cells. The day 14 images are representative of 3F and low-expressing 4F reprogramming to highlight the appearance of ESC-like colonies. Scale bars, 200 μm. **d**, Nanog expression in

day 30 colonies. High-expressing 4F cells (1B) exhibit the F-class cell morphology. Scale bars, 100 μm. **e**, Principal component analysis of quantitative RT–PCR values (32 genes). 3F, 4F low and 4F high cell lines are described in **a** and **b**. Cell-state landmarks are F-class clones 1 and 5 (red squares), and C-class clones 5, 10 and 23 (blue squares). **f**, Quantitative RT–PCR analysis of low-expressing 4F (Col1a1, grey line) and high-expressing 4F (2° 1B, black line) reprogramming, $n = 1$. **g**, Schematic model of proposed cell reprogramming routes.
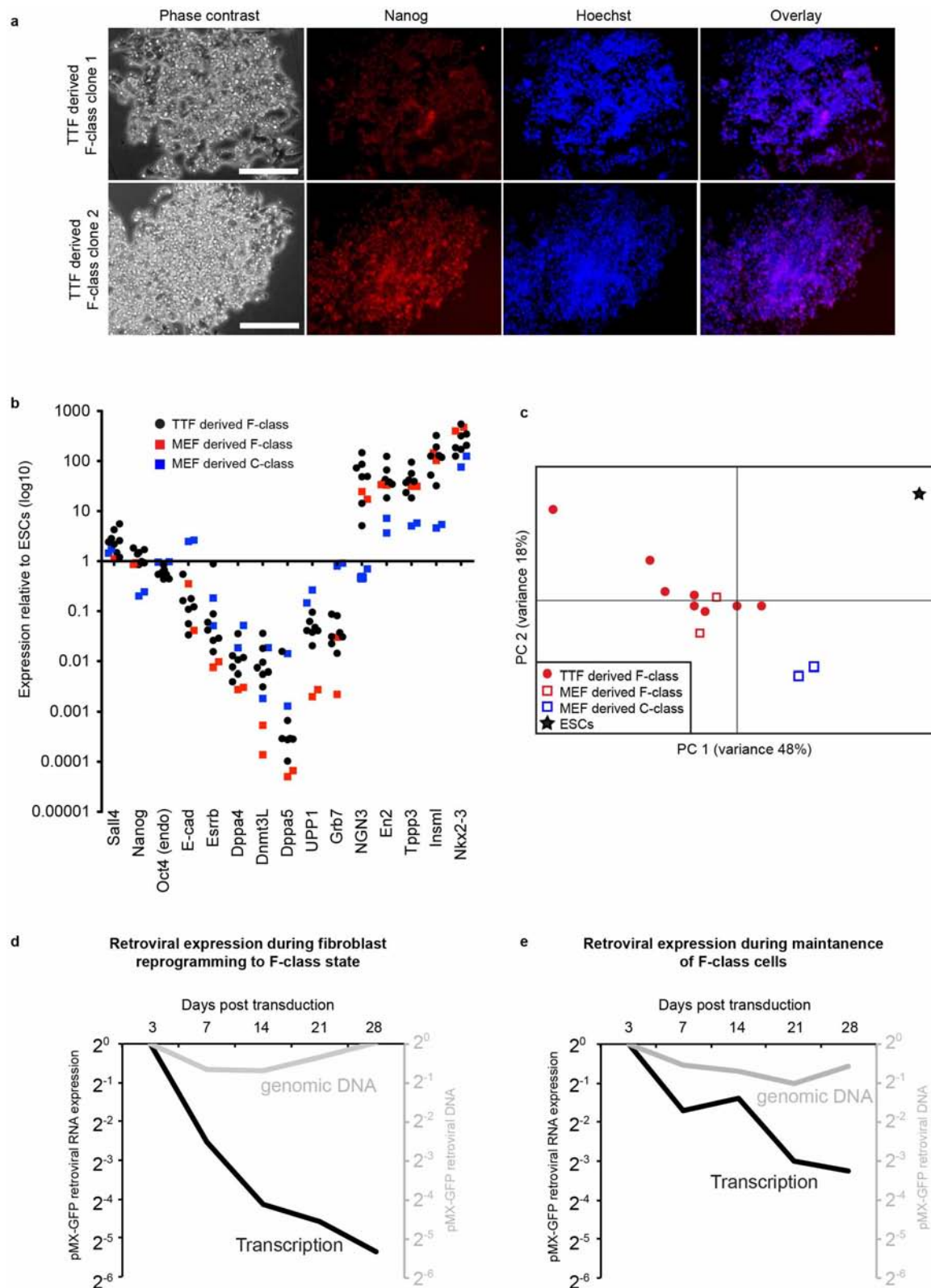
Extended Data Figure 7 | Adult tail tip derived F-class cells. a, Tail-tip fibroblast-derived F-class cells. Scale bars, 200 μm. **b**, Quantitative RT–PCR analysis of gene expression (day 25 of transgene expression) in clonal tail tip fibroblast reprogrammed cell lines $n = 7$ biological replicates. **c**, Principal component analysis of gene expression profile (quantitative RT–PCR, 32 genes). **d**, Retroviral silencing during transposon mediated reprogramming to F-class state. Quantitative PCR analysis of retroviral copy number (genomic DNA levels) and RNA transcription ($n = 3$ technical replicates from one experiment). **e**, Retroviral silencing in established F-class cells ($n = 3$ technical replicates from one experiment).
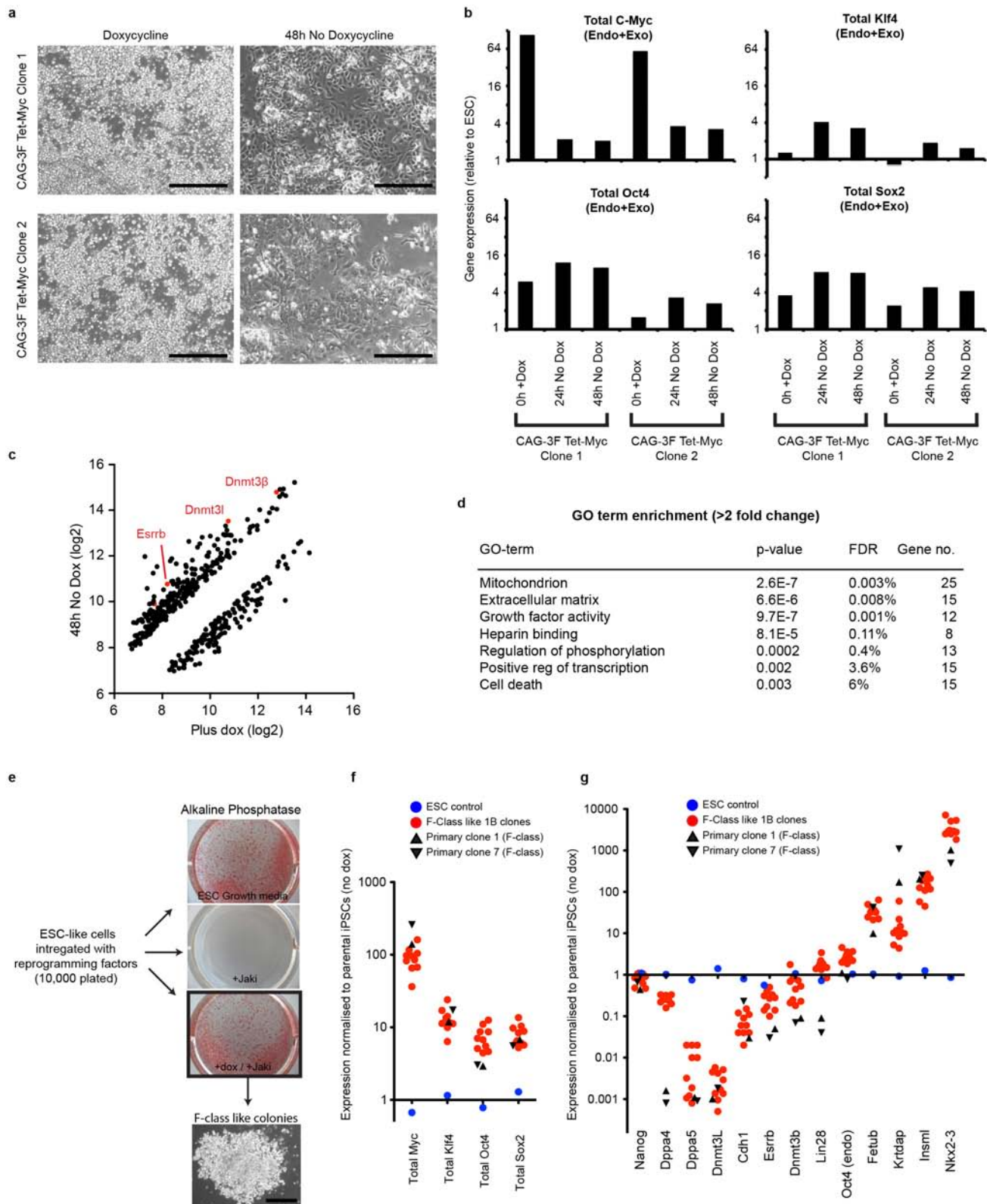
Extended Data Figure 8 | Requirement of four reprogramming factors.
a, Phase contrast images of F-class cells (CAG-3F + tetO Myc cells). Scale bars, 200 μm. **b**, Quantitative RT–PCR analysis of reprogramming factor expression, two independent cell lines (data are from $n = 2$ biological replicates with 3 technical replicates each). **c**, Genes exhibiting >twofold change upon doxycycline removal (Illumina BeadArray, two independent clones). **d**, Gene ontology term enrichment of differential gene expression. **e**, Reprogramming factor expression was activated in ESC-like cells (1B primary iPS cell line). **f**, Quantitative RT–PCR of reprogramming factor expression in cell lines ($n = 10$) established from F-class colonies picked in **e**. **g**, Quantitative RT–PCR expression of ESC and F-class gene identifiers in cell lines ($n = 10$) established from F-class colonies picked in **e**.
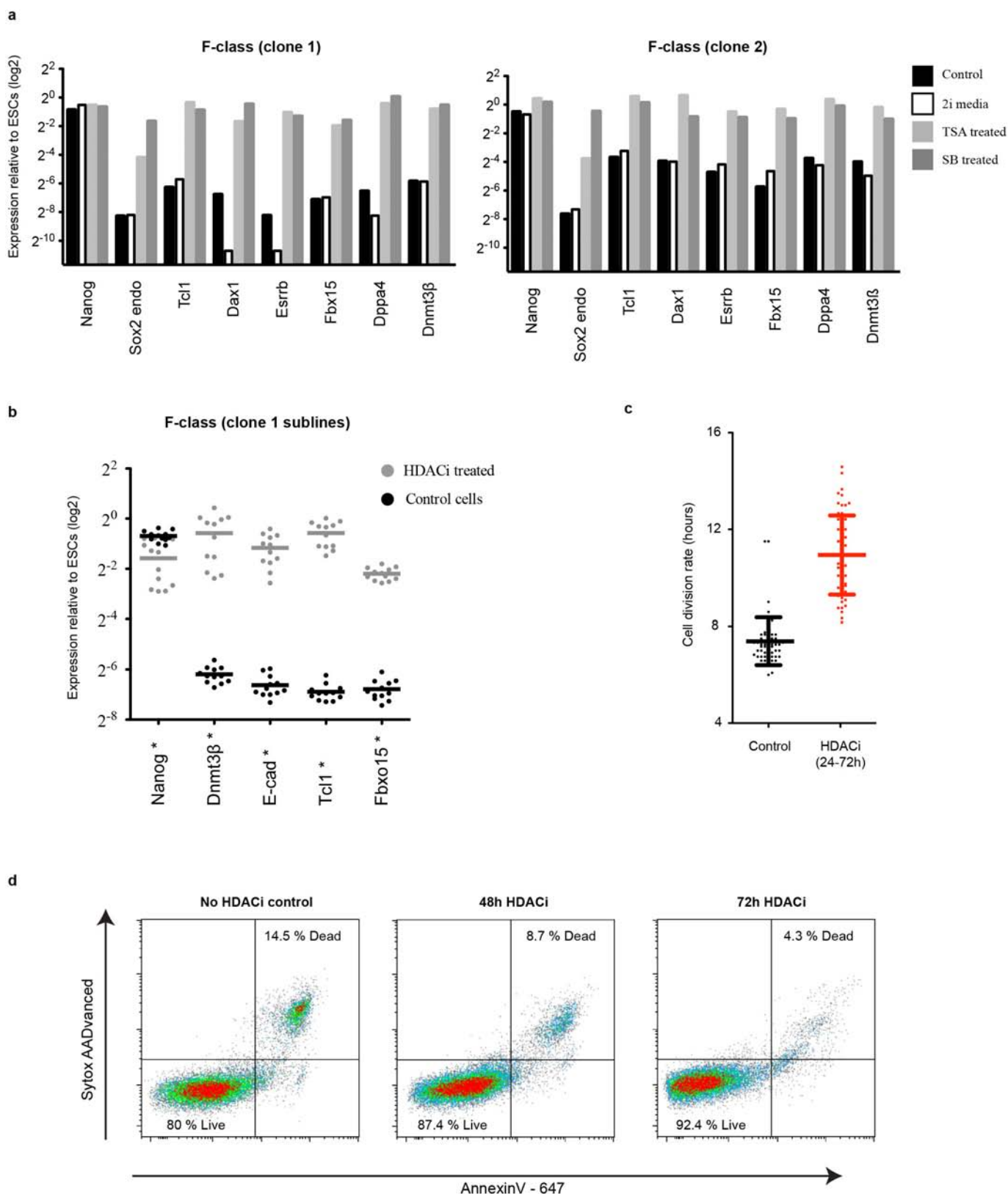
Extended Data Figure 9 | HDACi-induced transition to ESC-like state.
a, Quantitative RT–PCR of gene expression in F-class cells (day 30, doxycycline-supplemented) that were either maintained in 2i media or exposed to HDAC inhibitors for 10 days. Two cell lines representative of six F-class lines. Data are from $n = 3$ technical replicates from one experiment.

b, Quantitative RT–PCR of F-class (clone 1) sub-lines ($n = 12$) treated with 10 nM trichostatin A for 6 days. Line denotes average. **c**, Cell division rate of HDACi-treated cells, as determined directly by time-lapse analysis. **d**, Flow cytometric analysis of cell viability upon HDACi treatment (10 nM trichostatin A).
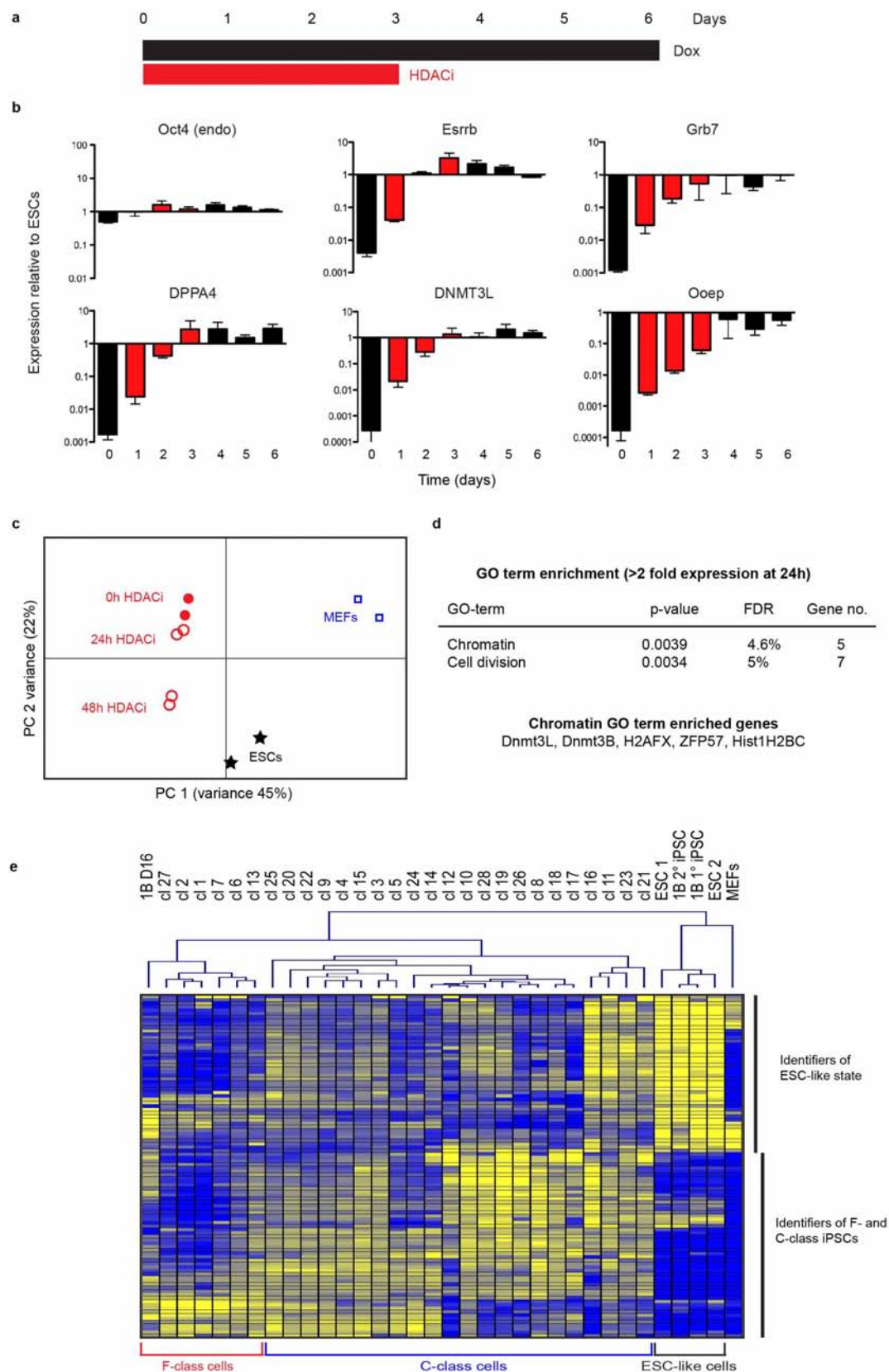
Extended Data Figure 10 | Temporal effect of HDACi. a, Schematic representation of HDACi treatment. **b**, Quantitative RT–PCR analysis of gene expression during HDACi treatment. Red bars depict time points of HDACi exposure (10 nM TSA). Data are from $n = 3$ biological replicates with 3 technical replicates each (average ± s.d.). **c**, Principal component analysis of gene expression (Illumina BeadArray). **d**, Gene ontology term enrichment analysis of genes during HDACi treatment. **e**, Unsupervised hierarchical clustering of gene expression (Illumina BeadArray) corresponding to primary reprogrammed clones after 16 days of transgene expression and day 16 cells from the 1B secondary reprogramming system (1BD16).

ARTICLE

Genome-wide characterization of the routes to pluripotency

Samer M. I. Hussein[1,*], Mira C. Puri[1,2,*], Peter D. Tonge[1,*], Marco Benevento[3,4], Andrew J. Corso[1,5], Jennifer L. Clancy[6], Rowland Mosbergen[7], Mira Li[1], Dong-Sung Lee[8,9], Nicole Cloonan[10,†], David L. A. Wood[10], Javier Munoz[3,4,†], Robert Middleton[11], Othmar Korn[7], Hardip R. Patel[6,12], Carl A. White[13,14], Jong-Yeon Shin[8,15], Maely E. Gauthier[10], Kim-Anh Lê Cao[10,†], Jong-Il Kim[8,9], Jessica C. Mar[16], Nika Shakiba[13], William Ritchie[11], John E. J. Rasko[11,17], Sean M. Grimmond[10], Peter W. Zandstra[13,14], Christine A. Wells[7,18], Thomas Preiss[6,19], Jeong-Sun Seo[8,9,15], Albert J. R. Heck[3,4], Ian M. Rogers[1,20,21] & Andras Nagy[1,5,21]

Somatic cell reprogramming to a pluripotent state continues to challenge many of our assumptions about cellular specification, and despite major efforts, we lack a complete molecular characterization of the reprograming process. To address this gap in knowledge, we generated extensive transcriptomic, epigenomic and proteomic data sets describing the reprogramming routes leading from mouse embryonic fibroblasts to induced pluripotency. Through integrative analysis, we reveal that cells transition through distinct gene expression and epigenetic signatures and bifurcate towards reprogramming transgene-dependent and -independent stable pluripotent states. Early transcriptional events, driven by high levels of reprogramming transcription factor expression, are associated with widespread loss of histone H3 lysine 27 (H3K27me3) trimethylation, representing a general opening of the chromatin state. Maintenance of high transgene levels leads to re-acquisition of H3K27me3 and a stable pluripotent state that is alternative to the embryonic stem cell (ESC)-like fate. Lowering transgene levels at an intermediate phase, however, guides the process to the acquisition of ESC-like chromatin and DNA methylation signature. Our data provide a comprehensive molecular description of the reprogramming routes and is accessible through the Project Grandiose portal at http://www.stemformatics.org.

Forced expression of four transcription factors—Oct4 (also called Pou5f1), Sox2, Klf4 and Myc (OSKM)[1]—induces molecular changes in somatic cells, which lead to pluripotency. The exogenous expression of these transcription factors perturbs the transcriptional network of the initial somatic cell. In response, the cells process intrinsic and extrinsic cues to remodel chromatin and reach a new epigenetic state. The newly established molecular profile of the cells is similar to embryonic stem cells, thus conferring upon the cells an ESC-like pluripotent state[2,3]. In the accompanying paper[4] we demonstrate that this is not the only induced pluripotent stem cell (iPSC) outcome, as we have characterized a novel category of steady-state pluripotent cells, named F-class after the fuzzy appearance of cell colonies in culture.

Despite rapid progression of our understanding of the reprogramming process, the cascade of molecular events defining the cellular outcomes are not well understood due to the low frequency of reprogramming to an ESC-like state. Without a significant increase in efficiency, the early, seemingly stochastic events of reprogramming are difficult to study. To overcome this limitation, secondary reprogramming systems provide sufficient number of cells[5–9]. Here we use iPSC-derived differentiated

cells containing doxycycline-inducible reprogramming transgenes[5,8,10]. Secondary reprogramming studies provide evidence that reprogramming is a multistep process in which iPSCs are reached via transitions through defined transcriptional and chromatin states[5–7,11]. However, the gene expression networks and their epigenetic basis in intermediate reprogramming states have not been defined in detail, and much is still unknown about the different outcomes to reprogramming. A molecular understanding of the different cell states that reprogramming generates provides a foundation for better control over the process.

Our generation and integration of multiple 'omic' profiles characterized two reprogramming routes at the highest resolution. The same cell collections were subjected to next-generation sequencing to determine the methylome (genome-wide CpG methylation), the transcriptome (short and long RNAs) and the chromatin marks (ChIP-sequencing for H3K4me3, H3K27me3 and H3K36me3 marks). In addition, we performed quantitative mass spectrometry to establish global and cell-surface proteomes.

We reveal that the early loss of H3K27me3 is associated with the acquisition of a transient open/primed chromatin state. Subsequent

reacquisition of H3K27me3 and an absence of DNA demethylation during sustained high transgene expression facilitate reprogramming towards the F-class pluripotent cell state. The gain of H3K4me3 chromatin marks is biphasic, with an early phase activating a subset of ESC-associated genes and a later phase reflecting reprogramming to an ESC-like state. The late-phase gain of H3K4me3 is accompanied by DNA demethylation leading to gene activation and stabilization of the ESC-like cell state. An integrative analysis of all the platforms further enabled us to refine the cohort of ESC-associated non-coding RNAs, to establish their epigenetic regulation and to reveal novel non-coding and protein-coding genes.

Experimental design and platforms

We took advantage of our *piggyBac* transposon-mediated, 1B[4,10] iPSC line-based, doxycycline-inducible secondary reprogramming system and modelled the reprogramming routes that lead to F-class and ESC-like iPSCs (Fig. 1a). Through gene expression analysis we demonstrate that sustained high-doxycycline reprogramming of the 1B secondary mouse embryonic fibroblast (MEF) generates F-class cells (Fig. 1a, D16H and D18H samples; where D indicates day and H indicates high doxycycline) that resemble the primary F-class cell lines of Tonge *et al.*[4]. However, lowering the doxycycline concentration after day 8 (D8H) facilitated reprogramming to the ESC-like state (Fig. 1a and Extended Data Fig. 1a–c), resulting in samples D21L, D21∅ (L, low; ∅, zero doxycycline) and secondary iPSCs. The secondary PSCs together with the primary and genetically related Rosa26-rtTA knock-in ESCs[12] (Fig. 1a) represent ESC-like cells, with the ability to contribute to embryonic development (Extended Data Fig. 1d, e). High-doxycycline samples maintained high levels of transgene expression (Extended Data Fig. 1f), demonstrating that the transgenes are not silenced as reprogramming progresses.

We performed 'multiple-omic' analyses on the samples taken from the reprogramming process (Fig. 1b and Extended Data Fig. 2) at the time points indicated in Fig. 1a (details in Methods). Normalized, curated

'omic' data are accessible through the Stemformatics platform (http://www.stemformatics.org), which provides analysis tools and facilitates a locus-centric visualization of 'omic' platforms on the UCSC genome browser[13] (Fig. 1b and Extended Data Fig. 2).

Multi-omic characterization of cell states

Pearson correlation distance analysis of long RNAs and microRNAs (miRNAs)[14] segregated the cell samples into six distinct categories: (1) secondary MEF; (2) D2H; (3) D5H–D11H; (4) D16H/D18H/D16L (F-class cell state); (5) D21L/D21∅; and (6) secondary iPSCs/ECSs/primary iPSCs (ESC-like state) (Fig. 2a and Extended Data Fig. 3a). Sample grouping was further supported by principal component analyses (PCA) of the global proteomic data[15] (Fig. 2b). PCAs of all platforms revealed that elevated transgene expression (high-doxycycline) drove secondary MEFs to a cell type distinct from the ESC-like state (Fig. 2b, red arrow, and Extended Data Fig. 3b–d). The reduction of doxycycline concentration at day 8 gave rise to samples D21L and D21∅ whose 'omic' profiles more closely resemble those of ESC-like pluripotency.

Upon addition of doxycycline to secondary MEFs, there was an immediate and dramatic change within each platform, except for the methylome (Fig. 2b and Extended Data Fig. 3b–d). The delayed kinetics of methylome remodelling are consistent with previous reports that suggest DNA methylation patterns are reset late in reprogramming[6,16,17]. To temporally define the transcriptional response to reprogramming factor expression we used Shannon entropy-based analysis[18] of protein-coding gene expression (Fig. 2c and Extended Data Fig. 4a). Gene ontology (GO) analysis of stage-specific gene expression identified enrichment for cell adhesion molecules and ectoderm development within the immediate responder genes (D2H) (Supplementary Table 1). Similarly, intermediate stage (D5H–D8H) genes were enriched for cell-adhesion molecules and immune response genes. The upregulation of genes associated with epithelialization is consistent with mesenchymal–epithelial transition (MET) being an early event of reprogramming[5,19]. Notably, GO analysis of F-class stage-specific genes revealed that they were enriched for 'cell fate commitment' and 'neuron development', suggesting that these cells are amenable to differentiation (Fig. 2d). Moreover, during high-doxycycline reprogramming, some ESC pluripotency genes including *Nanog* and *Sall4* were rapidly upregulated (early ESC genes) and remained expressed in F-class cells, highlighting similarities in the pluripotency network between the F-class and ESC-like states (Fig. 2e). However, a number of cell adhesion genes that are pluripotency related (*Cdh1* and *Epcam*) were suppressed in the high-doxycycline F-class cells, in comparison to ESC and iPSCs.

By comparing other transcriptome data sets to ours we found that at no point during reprogramming to the F-class state do the cells represent established pluripotent stem cells (such as epiblast stem cells) or early post-implantation primary cells[20] (Extended Data Fig. 4b). One study described multiple transient cell states captured by cell sorting for two cell-surface markers, ICAM1 and CD44 (ref. 8); however, no F-class equivalent stable state was described (Extended Data Fig. 4c). In particular, the Nanog[+] F-class cells were negative for both markers (Extended Data Fig. 4d), and when compared to the ICAM[−]CD44[−]Nanog[+] cell population of ref. 8 there was poor expression correlation (Extended Data Fig. 4c). Transient cell types are also described in ref. 6, by sorting for SSEA1 to capture cells that reprogram to the ESC-like state. Within 2 days of reprogramming factor expression, our reprogramming system generated cells that were similar to their D3–D9 SSEA1[+] sorted samples (Extended Data Fig. 4e). This is also consistent with our earlier observation showing that more than 70% of the reprogramming 1B secondary cells were SSEA1-positive at day 6 of reprogramming[10]. Therefore, our population-based transgene expression rapidly initiates a global cell-state change towards pluripotency without the need for enrichment.

The integration of public data sets with data presented here provides additional evidence that a key determining factor for the F-class state is the reprogramming transgene expression level[4]. We mined the transcriptomic

Figure 1 | Multi-omics analysis of secondary reprogramming. a, Outline of secondary reprogramming and designation of collected cell samples. Full details found in Methods. 1°, primary; 2°, secondary. **b**, Schematic representation of the data tracks at the *Dnmt3l* locus (chr10: 77,519,500–77,526,500 mm9 assembly), as hosted by http://www.stemformatics.org. For proteomics, 13 Dnmt3l-related peptides were detected and represented as scaled bars according to their corresponding exon positions.

Figure 2 | Molecular characterization of cell states during secondary reprogramming.
a, Merged pairwise comparisons (Pearson correlation) for long RNA-seq (mRNA and lncRNA) and short RNA-seq (miRNA). Sample groupings assigned on the basis of correlation distances (see Extended Data Fig. 3a) are numbered and colour coded. **b**, Principal component analysis of global protein, for all proteins quantified by mass spectrometry. Red arrow represents high-doxycycline reprogramming trajectory to the F-class state (red cloud). Black dashed arrow follows the low-doxycycline trajectory to the ESC-like state (grey cloud). **c**, Heat maps show stage-specific protein-coding gene expressions belonging to five categories: immediate responder, intermediate, F-class, early ESC and iPSC/ESC genes. Gene names in red are upregulated during intermediate stages of reprogramming and downregulated in the F-class state. **d**, GO analysis results for genes enriched in F-class state. **e**, Transcriptional activity of early ESC-like genes. Red and orange lines highlight increased expression of *Nanog* and *Sall4* pluripotency regulators throughout reprogramming, in contrast to decreased expression of cell adhesion proteins *Epcam* (black line) and *Cdh1* (blue line) specifically in F-class cells. For gene lists associated with **c–e**, refer to Supplementary Table 1. FPKM, fragments per kilobase of transcript per million mapped reads.

data of ref. 7 and examined correlations with data from our samples. The study of ref. 7 investigated transgene-independent reprogrammed cells (SC cells) and those that remain transgene dependent (SI lines). At the transcriptional level, SI cell lines exhibit higher correlation to F-class cells than ESC-like cells (Extended Data Fig. 5a). The doxycycline-exposed SI clones possessed higher reprogramming transgene expression than SC clones (Extended Data Fig. 5b), demonstrating that maintenance of high transgene expression is not compatible with reprogramming to an ESC-like state but drives reprogramming to alternative cell states. PCA and hierarchical clustering revealed that all of our reprogramming groups, including intermediate samples and F-class cells, clustered separately from doxycycline-exposed SI and SC clones, whereas iPSCs from both studies clustered together (Extended Data Fig. 5c, d). This suggests a difference between the route to the F-class state compared to the path taken by SI and SC clones. This view is supported by higher expression of the adhesion molecules *Cdh1* and *Epcam* and some ESC-like markers such as *Icam1*, *Nr5a2* and *Mycl1* in doxycycline-treated SI clones versus F-class cells (Extended Data Fig. 5d). Moreover, F-class cells showed higher expression of developmental genes, such as *Isl1*, *Gli1* and *Kit* (Extended Data Fig. 5d and Supplementary Table 1).

Dynamics of chromatin remodelling

To investigate how gene expression dynamics are reflected in histone modification changes, we tracked H3K4me3 (activating), H3K27me3 (repressing) and H3K36me3 (marking transcriptional elongation[21]) at annotated loci. Besides local alterations, no net global change was observed in H3K4me3 marks during reprogramming (Fig. 3a and Extended

Data Fig. 6a). In contrast, we observed an initial global loss of H3K27me3 upon expression of the reprogramming factors, which reached the minimum at D8H, followed by a gradual increase during cell transition to both F-class and ESC-like states (Fig. 3a and Extended Data Fig. 6a).

We observed that during the phase of H3K27me3 global loss, the expression of the H3K27me3 demethylases *Kdm6a* (also called *Utx*), *Jhdm1d* (also called *Kdm7a*) and *Phf8* was upregulated steadily starting at D2H (Extended Data Fig. 6b). In contrast, expression of the PRC2 complex members (*Eed*, *Ezh2*, *Suz12*) that catalyse trimethylation of H3K27 (reviewed in ref. 22) remained very low until the H3K27me3 mark was reacquired after day 8 (Extended Data Fig. 6b). Previous studies demonstrated that perturbation of enzymes that modify methylation or demethylation of H3K27 influences reprogramming[23,24], supporting the view that the observed genome-wide change in H3K27me3 affects the reprogramming process.

We detected a lack of concordance between messenger RNA and protein levels for the *Kdm6a* gene within the ESC-like path of reprogramming (Extended Data Fig. 6b). We investigated whether intron retention[25] plays a role and found that the level of mRNA expression negatively correlated with the intron retention values (Supplementary Information) for 2,591 genes that showed intron retention (Extended Data Fig. 6c). In the case of *Kdm6a*, intron retention is high in F-class cells, and in parallel, transcription rapidly increased fourfold relative to secondary MEF levels. In the low-doxycycline samples and the ESC-like category there is threefold-lower mRNA expression but higher protein levels than in the F-class. This lack of concordance could be explained by the significantly lower intron retention values in the low-doxycycline
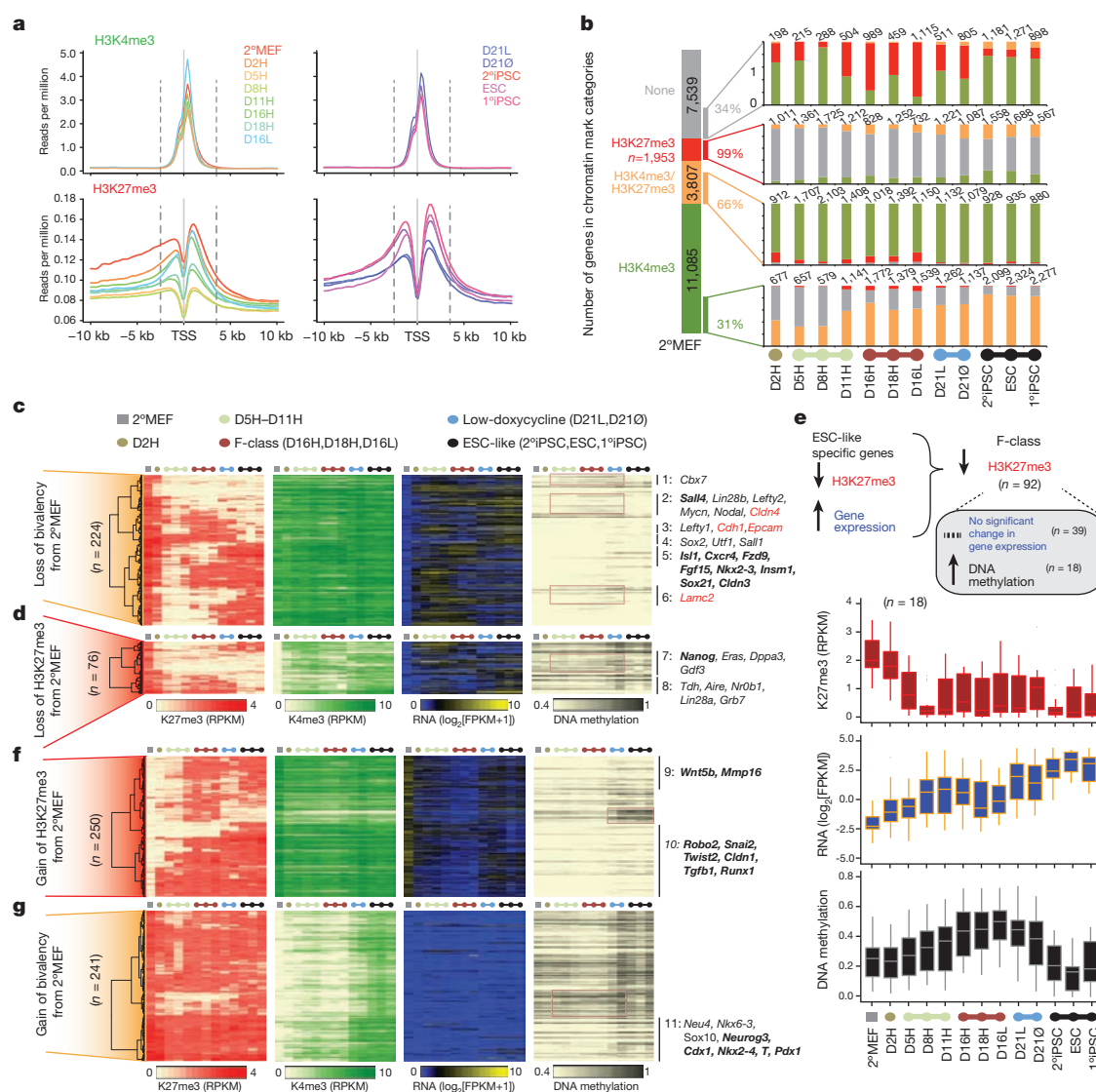
Figure 3 | Dynamic features of chromatin remodelling in reprogramming cell states. a, Read density profiles of H3K4me3 and H3K27me3 surrounding the TSS, as examined by ChIP-seq. **b**, Tracking of gene status with respect to histone mark content, relative to secondary MEF. **c**, Analysis of secondary MEF loci that are H3K4me3$^+$H3K27me3$^+$H3K36me3$^-$ and lose H3K27me3. Genes activated early during reprogramming and the ESC-like state, but silenced in the F-class state, are highlighted in red. Genes in bold type are activated in the F-class state. Red boxes on DNA methylation heat maps highlight gene clusters switching from a loss of H3K27me3 to an increase in DNA methylation. **d**, Analysis of transcriptionally silent (H3K36me3$^-$) loci that are monovalent (H3K27me3$^+$) in secondary MEF and gain H3K4me3 during reprogramming. **e**, Analysis of loci that lost H3K27me3 in ESC/iPSCs and F-class cells, and were activated in ESCs/iPSCs but not in F-class. Box plots

show levels of H3K27me3 enrichment, expression and DNA methylation for 18 genes that lost H3K27me3 but gained DNA methylation in F-class cells (represented by red boxes on DNA methylation heat maps in **c** and **d**). n = the number of genes for each category. Box plots represent the median (band inside the box), first and third quartiles. Whiskers extend to 1.5 times the interquartile range (IQR). Data points outside 1.5 times IQR are outliers. **f**, Analysis of loci that were H3K4me3$^+$ in secondary MEFs and acquired H3K4me3$^+$H3K27me3$^+$H3K36me3$^-$ (bivalent) profile. **g**, Analysis of loci that were H3K27me3$^+$ in secondary MEFs and became bivalent (H3K4me3$^+$H3K27me3$^+$H3K36me3$^-$) during reprogramming. For gene lists associated with **c**, **d**, **f**, **g** refer to Supplementary Table 2. RPKM, reads per kilobase of transcript per million mapped reads.

samples (Extended Data Fig. 6d). This is consistent with the previous finding that intron retention can counteract sudden changes in transcription[25]. In the case of ESC-like cells the decrease in intron retention provides more protein-coding RNAs for translation. When we considered genes whose chromatin status was unchanged at different stages of reprogramming, yet were differentially repressed, we observed a significant negative correlation between intron retention and RNA expression only in early reprogramming (from secondary MEF to D11H) (Extended Data Fig. 6e). This may suggest that during early reprogramming intron retention serves as a regulatory mechanism in the absence of H3K27me3 for a subset of genes.

The global reduction of H3K27me3 suggested a steady loss of heterochromatin up to D8H, followed by rebuilding of the H3K27me3 marks

in both routes of reprogramming. To obtain further evidence for this, we focused on transposable element expression, since transposable element silencing is linked to heterochromatin formation[26] and thus is implicated in chromatin organization[27,28] and gene regulation[29]. We examined four groups of transposable elements and observed that the number of expressed short interspersed elements (SINEs), long interspersed elements (LINEs) and DNA transposons was high during reprogramming factor expression, with the highest numbers consistently observed at D8H (Extended Data Fig. 6f). A similar pattern of expression was observed for L1 (LINE subfamily), the most actively expressed transposable element family in our data set (data not shown), and B2 (SINE subfamily) elements. Thus, the D8H peak in transposable element expression further supported that reprogramming led

to a transient opening of the chromatin, peaking at day 8 (D8H) of the process.

Given the depletion of H3K27me3 during the early phase of reprogramming, we categorized annotated genes with respect to associated chromatin marks observed in secondary MEFs: H3K4me3$^+$ monovalent, H3K27me3$^+$ monovalent, double positive (H3K4me3$^+$H3K27me3$^+$) and no mark (as shown in Fig. 3b). Global loss of H3K27me3 occurred in both H3K27me3$^+$ monovalent and H3K4me3$^+$H3K27me3$^+$ double-positive categories (see Extended Data Fig. 7 for a more detailed tracking of chromatin mark changes between samples). Ninety-nine per cent of secondary MEF H3K27me3$^+$ monovalent marks ($n = 1,953$) were lost by day 2, with most (>80%) remodelling to a no-mark status. In parallel, 66% of the H3K4me3$^+$H3K27me3$^+$ double-positive marks lost H3K27me3, to become H3K4me3$^+$ monovalent (Fig. 3b and Extended Data Fig. 7a, b). In contrast to the global loss of H3K27me3 marks, most secondary MEF H3K4me3$^+$ loci (69%) remained unchanged during reprogramming (Fig. 3b). A significant number of the remodelled secondary MEF H3K4me3$^+$ loci acquired H3K27me3 to become H3K4me3$^+$H3K27me3$^+$ double positive (Extended Data Fig. 7c). This occurred in both reprogramming routes to the F-class and the ESC-like states, with a higher proportion in the latter (Extended Data Fig. 7c).

Heterogeneity versus true bivalency

True bivalent (H3K4me3$^+$H3K27me3$^+$) primed loci are transcriptionally repressed[30–32]. Since H3K36me3 enrichment in the gene 'body' is a feature of expressed genes, bivalent loci are not expected to possess this histone modification. A significant proportion of H3K4me3$^+$ H3K27me3$^+$ double-positive loci, however, were also enriched for H3K36me3. RNA-seq data revealed that 75% of all triple-positive occurrences ($n = 4,137$, combined from all samples) were transcriptionally active, while this was true for only 24% of the H3K4me3$^+$ H3K27me3$^+$H3K36me3$^-$ loci ($n = 6,116$) (see Extended Data Fig. 8a for transcriptional threshold). These observations indicated that the majority of triple-positive loci represented a heterogeneous cell population: one subpopulation that is transcriptionally repressed (H3K4me3$^-$ H3K27me3$^+$H3K36me3$^-$ or H3K4me3$^+$H3K27me3$^+$H3K36me3$^-$) and the other transcriptionally active (H3K4me3$^+$H3K27me3$^-$ H3K36me3$^+$)[33–35] (see ref. 36 for a review). We examined the cell surface proteome and identified two proteins (CD24 or CD73) whose loci were triple positive in certain samples, suggesting population heterogeneity at the single-cell level. In these samples, flow cytometry indicated that protein expression was heterogeneous in the population. However, when H3K27me3 was lacking in a sample (H3K4me3$^+$H3K27me3$^-$ H3K36me3$^+$) the population homogeneously expressed CD24 and CD73 (Extended Data Fig. 8b). This indicated that the number of triple-positive loci can be used to estimate heterogeneity. Using these criteria, we found that high expression of the reprogramming factors has a unifying effect on the cell population; the heterogeneity was decreased (Extended Data Fig. 8c). The loss of triple-positive loci was primarily due to loss of H3K27me3, with loci remodelling to an active H3K4me3$^+$ H3K27me3$^-$H3K36me3$^{+/-}$ state (Extended Data Fig. 8d).

The poised H3K4me3$^+$H3K27me3$^+$ bivalent status of developmentally regulated loci is considered to be a hallmark of the epigenetic landscape of pluripotent ESCs and iPSCs[1,30,31]. Regarding true bivalency, we only considered H3K4me3$^+$H3K27me3$^+$H3K36me3$^-$ loci whose expression was below a defined expression threshold as determined by H3K4me3$^+$ active (H3K27me3$^-$H3K36me3$^+$) loci (Extended Data Fig. 8a). In parallel with global depletion of H3K27me3$^+$, the number of bivalent loci was reduced during the early phase of reprogramming, D2H–D8H (Extended Data Fig. 8e). However, we observed that concomitant with the reduced number of bivalent loci in early reprogramming was the acquisition of new bivalent domains, which continued throughout reprogramming, reaching the highest level of new bivalency in the ESC-like state (Extended Data Fig. 8f).

Interestingly, a number of secondary MEF bivalent marks were retained within all samples (Extended Data Fig. 8e). These were enriched

for the GO terms associated with cell-fate specification and developmental processes (Supplementary Table 2), which is consistent with previous reports[30,31]. Indeed, our meta-analysis of ChIP-seq data from refs 6 and 31 demonstrated that most bivalent domains in the MEFs from those studies are shared with 1B secondary MEFs (86% and 92%, respectively) as are those for ESC/iPSC-like pluripotent cells, indicating that the start and end points of reprogramming entail bivalency at highly overlapping loci (Extended Data Fig. 8g, h). However, comparison of 1B ChIP-seq data with that of intermediate reprogramming samples from ref. 6 demonstrated distinct differences in the temporal acquisition of bivalency between these reprogramming systems. Early reprogramming cells did not undergo a loss of bivalency[6] as seen with 1B secondary system, but instead acquired the majority of ESC/iPSC bivalent loci as early as day 3 (Extended Data Fig. 8i).

The rapid acquisition of bivalent loci seen with the SSEA1-positive cells of ref. 6 is consistent with them reprogramming primarily towards the ESC-like state. In contrast, 1B reprogramming cells rapidly acquired an open chromatin state in the early time points. From that state, the cells are more amenable to following multiple routes depending on the level of continued transgene expression (that is, high transgene expression drives the cells towards F-class cells, and low transgene levels towards an ESC-like state).

Gene regulation by H3K27me3

We questioned whether the loss of true bivalency controls expression of the stage-specific protein-coding genes. Bivalent loci that lost H3K27me3 by day 8 were then H3K4me3$^+$ monovalent (Fig. 3c), suggesting that they were primed for expression. Within this cohort were genes associated with cell adhesion (*Cldn4*) and pluripotency (*Lefty2*, *Lin28b*) (Fig. 3c, clusters 1, 2 and 3). The F-class state retained a number of these 'primed' genes (that is, *Sall4*); however, many genes became silenced by acquisition of H3K27me3 and/or DNA methylation (clusters 1, 2 and 3). Loci that reacquired H3K27me3 in the ESC-like state (H3K27me3$^+$H3K4me3$^+$) but did not possess H3K27me3 in the F-class cell state (Fig. 3c, cluster 5) were enriched for neural and cell migration genes such as *Nkx2-3*, *Isl1* and *Cxcr4* (Supplementary Table 2).

Next we investigated the transcriptional consequences of H3K27me3 loss at monovalent loci (H3K4me3$^-$H3K27me3$^+$) and found that *Nanog* and other pluripotency-associated genes were among them (Fig. 3d, cluster 7). However, not all of these genes were expressed in F-class cells despite showing loss of H3K27me3. This prompted us to investigate CpG methylation surrounding the transcriptional start site (TSS). We expanded this analysis to the bivalent loci that had also lost H3K27me3 and found that within both of these cohorts of genes were loci that had clearly gained DNA-methylation-based gene suppression during reprogramming (Fig. 3c, d, clusters 1, 2, 6, 7). Ninety-two ESC/ iPSC-specific genes lost H3K27me3 in F-class cells (Fig. 3e). Thirty-nine of these genes did not change in gene expression, while eighteen showed an increase in DNA methylation relative to MEFs. This may account for the low expression of these ESC-specific genes in the F-class state and suggests that continuously high transgene expression leads to the establishment of a DNA-methylation-based suppression of a subset of ESC/iPSC-specific genes. In an accompanying paper, it was observed that during reprogramming to the F-class state CpG methylation increased at or near binding sites of assessed transcription factors[37].

Several ESC-associated genes lost H3K27me3 and gained H3K4me3 during early reprogramming (Fig. 3d, cluster 8); however, several of these loci regained H3K27me3 in the F-class state, becoming bivalent. This switch from an H3K27me3$^+$ repressive mark to an H3K4me3$^+$ active mark defines one mechanism of transcriptional control during reprogramming, consistent with the loss of H3K27me3 conferring important gene regulatory function during reprogramming to pluripotent states.

A cohort of secondary MEF genes that became bivalent through acquisition of H3K27me3 (Fig. 3f) contained genes involved with the epithelial-to-mesenchymal transition (EMT) (Fig. 3f, clusters 9 and 10,

and Supplementary Table 2). This indicates that genes conferring a mesenchymal phenotype and MEF-specific cell adhesion are governed by changes in bivalency as observed by the gain of H3K27me3[+]. Conversely, H3K27me3[+] secondary MEF loci that became bivalent by addition of H3K4me3 (Fig. 3g, cluster 11) consisted of developmental regulators including lineage-specifying transcription factors (Supplementary Table 2). This primarily occurred in the low-doxycycline samples, generating poised developmental loci during reprogramming to the ESC-like state. A large proportion of ESC-associated bivalent loci were acquired from D5H, indicating that the modification of developmental genes to a poised status is an early event in reprogramming (Fig. 3g, cluster 11).

In summary, we have considered population heterogeneity from a chromatin perspective, to enable us to identify bivalent domains that were acquired throughout reprogramming, culminating in a distinct set of F-class and ESC-like specific bivalent marks. These bivalent domains repress gene expression to generate the divergent transcriptomes of the two cell states.

Building the new epigenome

Unlike the global H3K27me3 changes, we observed steady H3K4me3 mark levels during reprogramming (Fig. 3a and Extended Data Fig. 6a). However, changes in H3K4me3[+] sites, indicating active chromatin remodelling, did occur at the local level among a subset of loci ($n = 3,481$) (Fig. 3b). To determine the effect of H3K4me3 modulation on gene regulation, we examined genes that demonstrated differential H3K4me3 occupancy between secondary MEFs and the reprogramming samples (Fig. 4 and Supplementary Table 3). Sixty-three of seventy-nine secondary MEF H3K4me3[+] genes rapidly lost H3K4me3 within the first 5 days to reach an ESC-like chromatin status (H3K4me3[−] H3K27me3[−]) (Fig. 4a). This rapid loss of H3K4me3 coincided with the downregulation of transcription and gradual accumulation of DNA methylation during high-doxycycline reprogramming. Notably, F-class cells failed to reach the DNA methylation levels observed in ESC-like cells (Fig. 4a and ref. 37).

We next focused on loci that acquired H3K4me3 in the absence of H3K27me3 marks ($n = 108$). We found that 66% of these loci only acquired H3K4me3[+] during reprogramming to the ESC-like state (Fig. 4b and Extended Data Fig. 7f). This cohort of genes included some pluripotency network genes (*Dppa4*, *Dppa2*, *Dnmt3l* and *Dppa5a*), demonstrating that the upregulation of these ESC-associated genes is not connected with loss of H3K27me3. Notably, these genes showed reduced DNA methylation specifically in the ESC-like state (Fig. 4b and refs 4, 37). Intriguingly, more than 50% of these loci ($n = 39$) gained an H3K4me3 mark in low-doxycycline samples, but this change was not accompanied by DNA demethylation. At the TSS of these loci, H3K4me3 read density was inversely correlated with DNA methylation and was confined to regions that were hypomethylated (Fig. 4c). Moreover, H3K4me3 read density and peak width were significantly higher in secondary iPSCs compared to low-doxycycline, which maintained methylation levels similar to D8H and D16H (Fig. 4c, d). This indicated that the low-doxycycline samples had to overcome a DNA methylation barrier for progression to a fully reprogrammed ESC-like state. This is consistent with the findings of ref. 37, where H3K27me3 and H3K4me3 engagements were blocked in the presence of DNA methylation.

These results suggest that there are primarily two epigenetic determinants that influence whether cells reprogram to an F-class or ESC-like state. The first determinant is somatic cell inherited DNA methylation that must be removed to transition to an ESC-like state (see ref. 37). The second determinant is H3K27me3, which is acquired and maintained in the F-class state during high transgene expression, actively repressing genes associated with the ESC-like state.

Identification and regulation of lncRNA transcripts

We integrated chromatin and transcriptome data (Extended Data Fig. 9a, b and Supplementary Information) to identify long non-coding RNA
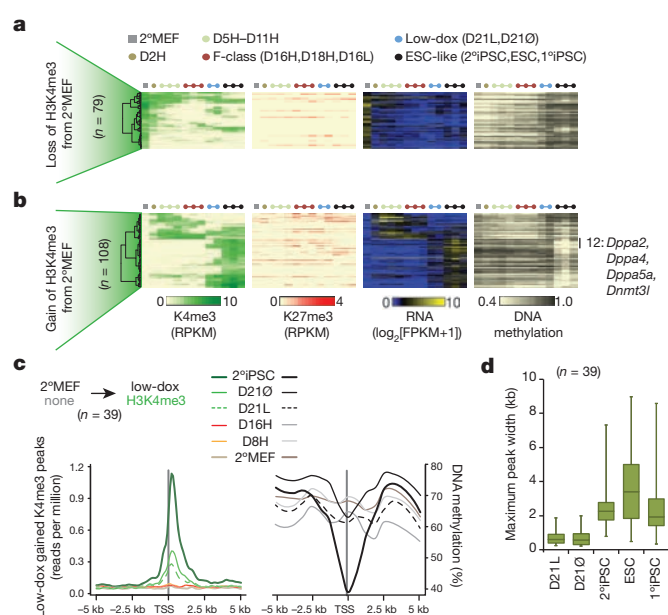
Figure 4 | H3K4me3 dynamics define cell states. a, Analysis of secondary MEF H3K4me3[+] loci that lack H3K4me3 in the ESC-like group. **b**, Analysis of secondary MEF H3K4me3[−] loci that gain H3K4me3 in the ESC-like group. **c**, H3K4me3 read density distribution and CpG methylation level profiles at TSS of loci that acquired H3K4me3 in low-doxycycline (low-dox) samples. **d**, Box plot of maximum H3K4me3 peak width of loci considered in **c**. n = the number of genes. Box plots represent the median (band inside the box), first and third quartiles. Whiskers extend to 1.5 times IQR. For gene lists associated with **a**, **b**, refer to Supplementary Table 3.

(lncRNA) transcripts that are regulated in a stage-specific manner during reprogramming. We identified 479 annotated and 767 unannotated lncRNAs that exhibited stage-specific expression during reprogramming (Fig. 5a, b and Supplementary Table 4). When unannotated lncRNA transcripts were categorized according to their genomic features (Extended Data Fig. 9c; see also Supplementary Information) we identified several that harboured miRNA clusters, enabling us to map miRNA TSSs for further studies on their transcriptional control (for analysis details refer to ref. 14). Many transcripts were multi-exonic (19%), a defining feature of lncRNAs[38] (Extended Data Fig. 9c). Among the unannotated lncRNA transcripts, several ($n = 96$) exhibited protein-coding potential (Extended Data Fig. 9d), prompting us to screen our proteome data for corresponding peptides. For three transcripts, we identified unique peptides that were abundant and differentially expressed during reprogramming. All three displayed a strong concordance between the protein and transcript expression throughout reprogramming (Extended Data Fig. 9e).

lncRNAs have been implicated in the regulation of the pluripotent state by their ability to interact with multiple chromatin regulatory proteins[39–41] and have been shown to prevent lineage-specific marker expression in ESCs[42]. We examined 226 ESC lncRNAs described by ref. 42 and found that 128 of these lncRNAs were expressed in our data set. Of these, 50 were differentially expressed during reprogramming, with 32 of them specifically expressed in the ESC-like group (Fig. 5c). A total of 48 of the 128 were found to be chromatin bound[42], of which 21 (lower red box) were from the 50 differentially expressed lncRNAs. A significantly high proportion (15 of 21) of differentially expressed chromatin-bound lncRNAs exhibited an association with PRC1 and PRC2 complexes ($P < 0.001$, binomial distribution test) compared to non-differentially expressed (9 of 27 ($P > 0.1$, binomial distribution test)). This is consistent with a suggested role of lncRNA[42] in determining chromatin state; in our case the dynamic change in the H3K27me3 chromatin marks during reprogramming (Fig. 3a).

To identify lncRNAs that have a conserved role during reprogramming we performed meta-analysis on raw RNA-seq data from published

a

Reprogramming stage-specific annotated lncRNA genes (n = 479)

RNA (log₂[FPKM+1]) — scale 0 to 8

b

Reprogramming stage-specific unannotated transcripts (n = 767)

RNA (log₂[FPKM+1]) — scale 0 to 5

c

ESC lncRNAs: ref. 42 (n = 226)

98 (43%)
78
50 Differentially expressed
18   32
ESC/iPSC-specific

Identified lncRNAs 128 (57%)

(n = 80)

Chromatin-bound lncRNAs

(n = 48)

RNA (log₂[FPKM+1]) — scale 0 to 8

d

Non-stage-specific lncRNA ■   Stage-specific lncRNA ●

Columns: PRC2, Ring1b, Jarid1b, Eset, Suv39h1, Cbx1, Setd8, Jarid1c, Cbx3, Yy1, Tip60/P400, Hdac1

H3K4me eraser / H3K27me reader / H3K27me writer
H3K9me writer / H3K9me reader / H4K20me writer
H3K4me eraser / H3K27me reader / H3K27me writer

Enrichment score — scale 0 to 5

e

- ■ 2°MEF
- ● D5H–D11H
- ● D2H
- ● F-class (D16H,D18H,D16L)
- ● Low-dox (D21L, D21∅)
- ● ESC-like (2°iPSC, ESC, 1°iPSC)

H3K27me3 modulation from 2°MEF (n = 131)

H3K4me3 modulation from 2°MEF (n = 141)

K27me3 (RPKM) 0 to 4 | K4me3 (RPKM) 0 to 10 | RNA (log₂[FPKM+1]) 0 to 6 | DNA methylation 0.4 to 0.8

Figure 5 | Expression and regulation of lncRNAs during reprogramming. a, b, Stage-specific expression of annotated (**a**) and unannotated lncRNAs (**b**). **c,** Expression heat maps of 128 lncRNAs out of 226 previously identified ESC lncRNAs[42], including 48 chromatin-bound lncRNAs (bottom heat map). Red boxes indicate differentially expressed lncRNA in 1B reprogramming. Pie chart indicates proportions of identified lncRNAs in 1B reprogramming and number of differentially expressed and ESC/iPSC specific transcripts. *n* = the number of genes represented in each category. **d,** Heat map illustrating enrichment values, as determined in ref. 42, of chromatin-associated lncRNAs (rows) for each chromatin-modifying enzyme (columns). Black and red dots indicate non-differential and differential expression in 1B reprogramming, respectively. **e,** Analysis of chromatin mark changes, DNA methylation and expression of differentially expressed lncRNAs in 1B reprogramming. *n* = the number of genes. For gene lists related to **a–e**, refer to Supplementary Table 4.

studies that used different reprogramming methods and cell types[7,8]. We determined that expression of stage-specific lncRNAs was transgene-level dependent (Extended Data Fig. 10a) and followed a sample correlation pattern as was observed with protein-coding genes (Extended Data Fig. 10a, b). Our analysis of lncRNA expression profiles of cell lines from ref. 7 was consistent with our comparison using protein-coding genes (Extended Data Fig. 5a) and provides further independent support for the existence of reprogramming stage-specific and

end-stage-specific lncRNA sets, suggesting their involvement in modulating reprogramming and consolidating stabilized pluripotent states.

We next considered differentially expressed lncRNA genes between ref. 7 and our system (Extended Data Fig. 10c), and examined the chromatin modification profiles associated with their expression (Fig. 5e and Extended Data Fig. 10d). The regulation of these transcripts followed chromatin change and DNA methylation patterns similar to ones we described above for protein-coding genes. For instance, early activation of lncRNA expression was primarily associated with loss of H3K27me3 and later expression changes were attributed to both DNA methylation and gain of H3K4me3 (Fig. 5e). We also observed that the dynamics of bivalency coincided with the regulation of lncRNA expression (Fig. 5e). These findings suggest that, similar to protein-coding developmental regulators, some non-coding RNAs acquire poised transcription status during reprogramming to pluripotency. Importantly, we have refined the cohort of ESC-associated lncRNAs and their expression profiles during reprogramming to either an F-class or ESC-like state (Extended Data Fig. 10c). Furthermore, we established that lncRNA epigenetic control is analogous to that of protein-coding genes.

Transcriptional paths to two pluripotent cell states

We tracked gene expression relative to D8H, the bifurcation point of the reprogramming paths and also the highest level of open chromatin (Fig. 3a and Extended Data Fig. 6a). We subdivided these genes according to whether they were upregulated, maintained or downregulated from secondary MEF to D8H. We then restricted the gene list to those that were differentially expressed from D8H between F-class and ESC-like states (Fig. 6a). These subdivisions were then grouped according to whether their expression increased or decreased from D8H towards the F-class or ESC-like state. Genes that were downregulated in F-class cells relative to D8H (groups 3, 6 and 9, Fig. 6a and Extended Data Fig. 10e) were enriched for cell adhesion components, consisting of cell–cell junctions, extracellular matrix proteins and cell-surface adhesion receptors (Supplementary Table 5). This observation could account for the morphological differences between F-class and ESC-like cells[4]. F-class-specific upregulated genes (group 4a) included replication-dependent histone genes, consistent with a requirement for increased nucleosome assembly in conditions of rapid proliferation. These genes had lower expression in ESC-like cells, which proliferate at a lower rate compared to F-class cells[4]. The accompanying global proteome analysis strengthens this finding since F-class cells are highly enriched for proteins associated with metabolism and cellular proliferation[15]. In contrast, a subset of pluripotency and early developmental patterning genes was significantly upregulated in the ESC-like state (groups 1b, 4b and 7b) relative to D8H (Fig. 6a, Extended Data Fig. 10f and Supplementary Table 5). These genes remained unchanged within the F-class (groups 2, 5 and 8). Among these were primed genes that were transiently activated (D5H–D11H) and subsequently repressed in F-class, yet active in low-doxycycline and the ESC-like state (Fig. 3c, d). This indicates the acquisition of chromatin-driven repression in this set of genes in F-class cells. Additionally, several loci in F-class cells showed a lack of enrichment of the H3K4me3 mark relative to the ESC-like state, in combination with an absence of ESC-like state-specific DNA demethylation, suggesting another mechanism of blocking transcriptional activity in F-class cells (Fig. 4b). Overall, these findings indicate that the epigenetic state of reprogramming cells diverges from D8H onwards to yield the F-class and ESC-like phenotypes. Transgene-driven chromatin-remodelling events hold F-class cells in a state of low cell adhesion, while DNA methylation maintains quiescence of many ESC-associated genes.

Concluding remarks

Reprogramming somatic cells to pluripotency entails the execution of a complex sequence of transcriptional and epigenetic events that result in an alteration of the cell state. We characterized two paths of reprogramming: one that gives rise to ESC-like cells and a second path to
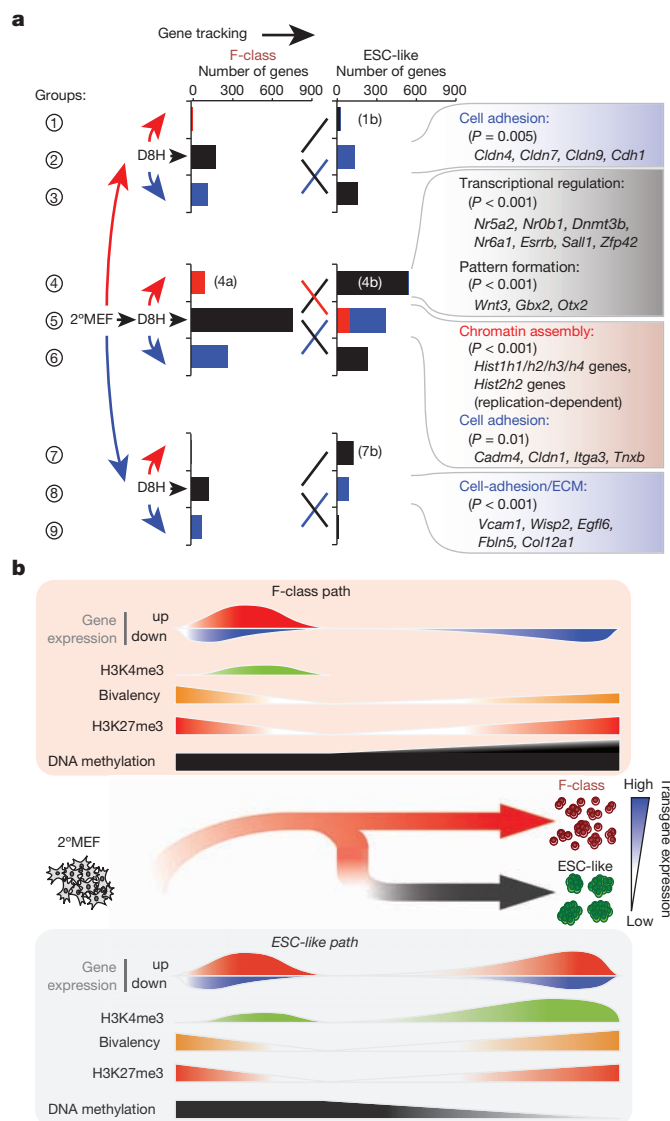
Figure 6 | Paths to F-class and ESC-like pluripotency. a, Schematic representation for tracking differentially expressed genes from secondary MEF to D8H, from D8H to F-class, from D8H to ESC-like cells, and their transition from F-class to ESC-like cells. Enriched GO terms and example genes are highlighted on the right for each group. For gene lists, full GO term analyses and *P* values associated with panel **a** refer to Supplementary Table 5. **b**, Summary illustration of presented data. Reprogramming cells follow two paths—F-class and ESC-like state paths—with specific patterns of gene expression, chromatin and DNA methylation changes that are dependent on transgene level change.

F-class[4] pluripotent cells (Fig. 6b). Our data demonstrate that during the first 8 days of induced cellular reprogramming, global patterns of H3K27me3 and H3K4me3 remain closely tied to transcriptional programs and cellular state. This finding suggests that activation of high-level OSKM expression transiently induces massive chromatin opening, caused by global loss of H3K27me3. Beyond day 8 of reprogramming factor expression, the suppressive H3K27me3 marks were gradually restored to levels observed in secondary MEFs and the ESC-like group. The ability to integrate and interrogate multiple 'omic' platforms has enabled us to discover novel protein-coding transcripts and characterize them with respect to epigenetic marks during reprogramming. Furthermore, we have refined the previously reported >200 ESC-associated lncRNAs[42] and identified the core lncRNAs that are associated with reprogramming to pluripotent states.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
3. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).
4. Tonge, P. D. *et al.* Divergent reprogramming routes lead to alternative stem-cell states. *Nature* http://dx.doi.org/10.1038/nature14047 (this issue).
5. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
6. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
7. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Stem Cells* **11,** 769–782 (2012).
8. O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* **499,** 88–91 (2013).
9. Nagy, A. Secondary cell reprogramming systems: as years go by. *Curr. Opin. Genet. Dev.* **23,** 534–539 (2013).
10. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).
11. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150,** 1209–1222 (2012).
12. Belteki, G. *et al.* Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res.* **33,** e51 (2005).
13. Wells, C. A. *et al.* Stemformatics: visualisation and sharing of stem cell gene expression. *Stem Cell Res.* **10,** 387–395 (2013).
14. Clancy, J. L. *et al.* Small RNA changes en route to distinct cellular states of induced pluripotency. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6522 (2014).
15. Benevento, M. *et al.* Proteome adaptation in cell reprogramming proceeds via distinct transcriptional networks. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6613 (2014).
16. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnol.* **28,** 848–855 (2010).
17. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biol.* **13,** 541–549 (2011).
18. Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6,** R33 (2005).
19. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
20. Kojima, Y. *et al.* The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14,** 107–120 (2014).
21. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128,** 707–719 (2007).
22. Simon, J. A. & Kingston, R. E. Occupying chromatin: polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49,** 808–824 (2013).
23. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488,** 409–413 (2012).
24. Pereira, C. F. *et al.* ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* **6,** 547–556 (2010).
25. Wong, J. J.-L. *et al.* Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154,** 583–595 (2013).
26. Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature Struct. Mol. Biol.* **20,** 332–338 (2013).
27. Tang, S.-J. Chromatin organization by repetitive elements (CORE): a genomic principle for the higher-order structure of chromosomes. *Genes* **2,** 502–515 (2011).
28. Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317,** 248–251 (2007).
29. Rebollo, R., Romanish, M. T. & Mager, D. L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46,** 21–42 (2012).
30. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
31. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
32. Jørgensen, H. F. *et al.* Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells. *Cell Cycle* **5,** 1411–1414 (2006).
33. Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151,** 181–193 (2012).
34. Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42,** 330–341 (2011).
35. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286,** 7983–7989 (2011).

36. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27,** 1318–1338 (2013).
37. Lee, D.-S. *et al.* DNA methylation as a reprogramming modulator: an epigenomic roadmap to induced pluripotency. *Nature Commun.* http://dx.doi.org/10.1038/ncomms6619 (2014).
38. Guttman, M. *et al. Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnol.* **28,** 503–510 (2010).
39. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
40. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106,** 11667–11672 (2009).
41. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
42. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477,** 295–300 (2011).

Supplementary Information is available in the online version of the paper.

Author Contributions S.M.I.H., M.C.P., P.D.T. and A.N. conceived, designed and carried out most of the experiments, interpreted results and wrote the manuscript. P.W.Z. contributed to study design. T.P., C. A. Wells, I.M.R., P.W.Z., C. A. White, N.S., A.J.C. and J.C.M. assisted with data interpretation and manuscript writing. M.L., S.M.I.H. and M.C.P. performed ChIP. M.C.P., S.M.I.H., N.C., O.K., D.L.A.W., M.E.G. and S.M.G. produced and analysed RNA-seq data. S.M.I.H., D.-S.L., M.C.P., J.-Y.S., J.-I.K. and J.-S.S. produced and analysed MethylC-seq and ChIP-seq data. J.E.J.R, W.R. and R.Mi. performed the IR analysis, interpretation and contributed to the manuscript writing. C. A. Wells, R.Mo., O.K., K.-A.L.C. and J.C.M. provided support for bioinformatics analyses and data visualization. M.B., J.M. and A.J.R.H. performed the LC-MS analysis and proteomic data analysis. H.R.P. mapped the miRNA Next Generation Sequencing (NGS) data and provided support for bioinformatics analyses and data visualization. J.L.C. and T.P. analysed and interpreted the miRNA NGS data. C.A.W. performed the CSC proteomics. C.A.W., N.S. and P.W.Z. analysed CSC proteome data.

METHODS

No statistical methods were used to predetermine sample size.

Cell culture and secondary reprogramming. ESCs and iPSCs were cultured in 5% CO_2 at 37 °C on irradiated MEFs in DMEM containing 15% FCS, leukaemia-inhibiting factor, penicillin/streptomycin, L-glutamine, nonessential amino acids, sodium pyruvate, and 2-mercaptoethanol. 1B primary iPS cells were aggregated with tetraploid host CD-1 embryos as described[43] (in compliance with Protocol 009 at the Toronto Centre for Phenogenomics) and MEFs were established from E13.5 embryos[43]. High doxycycline cell samples (1,500 ng ml^{-1} doxycycline) were collected at days 0, 2, 5, 8, 11, 16 and 18 (D2H, D5H, D8H, D11H, D16H, D18H). A subculture of the reprogramming cells was established from day 19 and cultured in the absence of doxycycline, to develop a factor-independent secondary iPS cell line by day 30. Low doxycycline samples were maintained from day 8 to day 14 cells in 5 ng ml^{-1} doxycycline. At day 14 the culture diverged to two groups, with one group of the cells collected at day 16 and day 21, from the low doxycycline concentrations of 5 ng ml^{-1} (D16L and D21L, respectively), and the other cultured until day 21 in the absence of doxycycline (D21Ø). ROSA26-rtTA-IRES-GFP mouse ES cells[12] and 1B primary iPSCs[10] were collected as controls. All cell lines tested negative for mycoplasma and other pathogens.

Long RNA sequencing and alignment. Cells were scraped, harvested in ice-cold PBS and stored in RNA-later (Ambion) at −80 °C. Total RNA for transcriptome sequencing was prepared using Qiagen total RNA purification kit followed by two rounds of on column DNase I treatment to remove contaminating DNA using the RNase-Free DNase set (Qiagen PN 79254) as per the manufacturer's protocol. The total RNA was then analysed using Agilent RNA 6000 Nano kit (PN 5067-1511) on the Agilent Bioanalyzer 2100 (PN G2939AA) to quantify yield, qualify integrity and confirm removal of DNA contamination.

Following DNase I treatment, 5 µg total RNA from each sample was depleted of Ribosomal RNA using the Ribo-ZerorRNA Removal kit (Epicentre PN RZH110424) as per manufacturer's instructions. The rRNA-depleted RNA was then run on an Agilent RNA 6000 Pico kit (PN 5067-1513) on the Agilent Bioanalyzer 2100 to confirm rRNA depletion. Sequencing libraries where generated from the rRNA-depleted RNA using the SOLiD Transcriptome Multiplexing kit (PN 4427046) from Applied Biosystems following the manufacturer's publication. Final libraries were quantified and qualified using Agilent High Sensitivity DNA kit (PN 5067-4626) on the Agilent Bioanalyzer 2100.

Sequencing libraries were subsequently pooled in equimolar ratios (four libraries per pool) and clonally amplified onto SOLiD Nanobeads. Clonal amplification was completed via emulsion PCR using the SOLiD EZ Bead System (PN 4448419, 4448418 and 4448420) coupled with SOLiD EZ Bead N200 amplification reagents (PN 4467267, 4457185, 4467281, 4467283, 4467282). Following emulsion, PCR clonally amplified Nanobeads were enriched using the SOLiD EZ Bead Enricher kits (4467276, 4444140, 4453073) before being deposited into a SOLiD 6-Lane Flow-Chip (PN 4461826) using the SOLiD Flowchip Deposition kit v2 (PN 4468081) as per the manufacturer's recommendations.

In total two flowchips were sequenced yielding a total of 8 lanes of data; with sequencing reads generated using the SOLiD 5500xl platform generating paired 75 bp forward and 35 bp reverse reads. To allow de-convolution of the pooled libraries a single 5 bp index read was generated. A total of 1,204,676,394 fragments (2,409,352,788 reads) were generated post de-convolution, ranging from 35,714,748 to 147,282,580 fragments per library (Supplementary Table 6).

Sequence mapping was performed using Applied Biosystems LifeScope v2.5 whole transcriptome (paired-end) analysis pipeline against the NCBIM37 (mm9) genome and exon-junction libraries constructed from the Ensembl v64 gene model. Briefly, this pipeline first removes potential contaminant reads by aligning to a filter set containing rRNA, tRNA, adaptor sequences and retrotransposon sequences. Following filtering, LifeScope then aligns all reads to the genome and F3 reads to the junction library. F5 reads are additionally aligned at a higher sensitivity to exonic sequences within insert size distance from the paired (F3) read alignment. For RNA-seq data sets from O'Malley et al.[8] (accession number E-MTAB-1654 in ArrayExpress) and Golipour et al.[7] (Gene Expression Omnibus (GEO) accession number GSE-42100), Tophat (version 2.0.6) was used to map reads against the NCBIM37 (mm9) genome (Ensembl v67). Read alignments were merged and disambiguated, and a single BAM (Binary Alignment Mapped) file output per library or sample was used. BAM files were then additionally filtered to remove reads with a mapping quality (MAPQ) <13, and all ribosomal and mitochondrial RNA reads. Alignments were assembled using Cufflinks (v2.1.1) using the –g parameter to construct a genome annotation file against the reference gene model (Ensembl v67) and to identify novel transcripts (refer to 'Long RNA sequencing analysis pipeline' below for details). *Long RNA sequencing analysis pipeline.* Run Lifescope: (i) align to filter set; (ii) align to genome; (iii) align to exon junction; (iv) choose alignments. (Refer to Supplementary Table 6 for read counts.) Remove reads <13 MAPQ. Remove chrMT reads.

Assemble with Cufflinks to create annotation file: (i) sequences assembled using −g parameter against Ensembl v67. FPKM values are then calculated as detailed below. FPKM values for each gene are used for all subsequent analyses.

Read count and differential gene expression. Raw read counts were obtained by mapping reads at the gene level using the Cufflinks assembled transcript annotation file with HTSeq-count tool from the Python package HTSeq, at http://www-huber.embl.de/users/anders/HTSeq/doc/count.html, using intersection-nonempty counting mode. EdgeR R-package (v3.0.8)[44] was then used to perform statistical analysis, samples were grouped as follows: group 1, secondary MEF; group 2, D2H; group 3, D5H–D11H; group 4, D16H–D18H, D16L; group 5, D21L and D21Ø; group 6, secondary iPSC, ESC and primary iPSC.

A common biological coefficient of variation (BCV) and dispersion (variance) was estimated for each grouping scenario based on a negative binomial distribution model. The estimated dispersion values were combined to obtain a final BCV value. This value was then incorporated into the final EdgeR analysis for differential gene expression, and the exact test for negative binomial distribution was used for statistics, as described in EdgeR user guide.

Identification and characterization of novel transcripts. Transcripts that did not overlap with Ensembl annotations were selected as candidate novel lncRNA genes. We only considered genes that fell under the following criteria: (1) length of 200 bp or more from cufflinks assembled transcriptome; (2) intergenic transcripts; (3) novel antisense transcripts; and (4) novel transcripts that overlapped intergenic miRNAs. Novel transcript overlapping annotated genes or novel isoforms as defined by Cufflinks output were not considered. We used Coding Potential Calculator (CPC)[45] to calculate the likelihood of any of these transcripts to be part of a coding protein sequence (that is, coding potential score). CPC accounts for quality and length of open reading frame, start and in-frame stop codons, and sequence homology with known protein-coding genes. Transcripts with a negative coding potential score were considered non-coding (Extended Data Fig. 9d). To identify multi-exonic novel transcripts, we relied on H3K4me3/H3K36me3 chromatin domains derived from our ChIP-seq data set, as previously described[41], to determine whether single transcripts with the same orientation as identified by Cufflinks were exons of a larger novel transcript. If these single transcripts were within an H3K4me3/H3K36me3 chromatin domain and showed a similar expression pattern, they were considered putative exons of one novel transcript.

TSS/gene promoter identification and FPKM calculation. To properly identify gene promoters and gene size for fragments per kilobase per million reads (FPKM) calculations, we first examined all possible annotated gene isoforms identified by Cufflinks and novel transcripts that passed the criteria described above, and restricted our analysis as follows:

(1) We divided every exon identified by Cufflinks assembled transcriptome into bins based on exon boundaries derived from all isoforms of a gene, identical to the method employed in DEXseq R package for differential exon usage[46]. We then mapped reads to these features using HTseq-count tool in intersection-nonempty counting mode. EdgeR R-package (v3.0.8)[44] was then used to normalize the data and calculate counts per million reads (CPM) values.

(2) For annotated isoforms, we only considered isoforms that showed at least 10 reads or a value of 0.2 CPM in the first exon bin in at least two samples, except when reads were only detected in secondary MEFs. Isoforms that failed to show expression in their corresponding first exon bin were filtered out.

(3) If multiple isoforms were detected based on the first exon bin strategy above, we examined the number of reads in the subsequent exon bin, sequentially scanning each exon bin of a gene for read count. Any exon bin that failed to have at least 10 reads or a value of 0.2 CPM was excluded along with its corresponding isoform. We followed this strategy until we identified the most abundantly expressed isoform per gene. We were able to robustly detect the proper isoform, with the exception of a few cases described below. The most abundant isoform size and TSS were used for FPKM calculations (FPKM = $1,000 \times$ CPM/(size of gene)) and subsequent analysis.

(4) The low number of genes, where this strategy failed, were genes with very low level expression. For such genes, we used Ensembl NCBI37/mm9 annotated mouse reference for TSS coordinates and calculated the gene size based on the sum of exon sizes where an expressed exon bin could be detected, and incorporated this into our FPKM calculations. In cases where a gene was not expressed in any of the samples, we used Ensembl NCBI37/mm9 annotated mouse reference for gene and TSS coordinates.

(5) For novel transcripts, we used annotations and gene sizes defined by Cufflinks *de novo* assembly, except in cases where we defined a novel transcript as multi-exonic. In these cases, the TSS of the first transcript (that is, first exon) within a novel multi-exonic gene was considered to be the start site, and the sum of transcript sizes as the final gene size.

Identification of stage-specific genes. To identify stage-specific genes, we used Shannon entropy modelling to compute a stage-specificity index for each gene, as

previously described[18,47]. Briefly, for each gene, a relative expression (Ri) value was calculated per sample and per grouping as described above and in Extended Data Fig. 3a, where Ri = (expression per sample or average expression per group)/sum of FPKM values in all the samples or all the groups. The entropy index score (Hi) across all samples or groups was calculated as $Hi = -1 \times \Sigma(Ri \times \log_2(Ri))$.

An entropy score close to 0 indicates high stage specificity, whereas a score closer to $\log_2$ of the total number of samples (13) or groups (6) indicates ubiquitous expression. As a threshold for selecting candidate genes for stage specificity, we used a tenth percentile cutoff (indicated by dashed lines in Extended Data Figs 4a and 9b) of the entropy scores distribution curve. Genes below this threshold were considered stage specific.

Analysis of repeat elements. We downloaded the RepeatMasker annotation file from the UCSC genome browser. We excluded repeats that overlapped Ensembl NCBI37/mm9 annotated mouse reference genes and considered repeats that are 200 bp or greater in length. Reads were mapped to these features using HTseq-count tool in Intersection-nonempty counting mode. EdgeR R-package (v3.0.8)[44] was then used to normalize the data and calculate CPM values. CPM values were then divided by the length of the repeats and multiplied by 1,000 to obtain FPKM values. Repeats with values >0.5 FPKM were considered expressed.

Calculation of transgene versus endogenous gene expression for *Oct4, Sox2, Klf4* and *Myc* factors. To obtain endogenous expression of the reprogramming factors, we followed two strategies. (1) We mapped reads to the 5′ UTR and 3′ UTR for *Sox2, Klf4* and *Myc* using these coordinates: (a) *Sox2*, 5′ UTR chr3:34548929–34549232, 3′ UTR chr3:34550301–34551414. (b) *Klf4*, 5′ UTR chr4:55544734–55545078, 3′ UTR chr4:55540033–55540942. (c) *Myc*, 5′ UTR chr15:61817049–61819045, 3′ UTR chr15:61821469–61821815.

(2) For *Pou5f1* (also known as *Oct4*), we identified a C/T single nucleotide polymorphism (SNP) at chr17 35643135 that differentiates between endogenous (C/G base pair) and exogenous (T/A base pair) expression. By mapping reads to the different polymorphisms, we quantified the relative levels of exogenous versus endogenous expression.

Library normalized read counts (CPM values) obtained from endogenous locations were further scaled for comparison to total CPM values. Scaling factors for each reprogramming factor were calculated from four samples (secondary MEF, ESC, secondary and primary iPSCs) as follows:

$$\text{Scaling factor} = \frac{\text{total expression of reprogramming factor in sample}}{\text{endogenous expression of reprogramming factor in sample}}$$

The scaling factor was averaged over the four samples and used to scale up endogenous CPM values for all samples. Exogenous reprogramming factor expression was determined as the difference between total and scaled endogenous expression.

Calculation of intron retention. Data were mapped to Ensembl assembly Mus_musculus.GRCm38.74. The same build was used to define gene structures. The intron retention ratio was calculated for each intron as depth of intron cover/(spliced reads + depth of intron cover).

Introns were listed as having significant intron retention if meeting the following conditions: (1) all samples had >5 reads correctly spliced across the intron; one sample had at least 20 reads; (2) one sample had reads covering >90% of non-excluded bases within the intron; (3) one sample had reads supporting continuation from exon into intron at both ends with minimum of 5 bp overhang; (4) to ensure readings were above any background, intron read depth was at least 25% greater than any neighbouring introns, or the neighbouring intron itself had been determined to have significant intron retention.

The mean of introns with significant intron retention within that gene was calculated for each gene. The following regions were excluded from intron retention analysis: (1) intronic regions that overlapped with exons, lncRNA and all other non-intron annotated features; (2) regions of poor mappability were excluded from statistics; (3) introns were excluded where a feature of opposite sense intersected; (4) introns were excluded if more than 30% of bases had been excluded or the length was less than 120 bp.

Introns contain numerous repeat and low-complexity regions to which software cannot uniquely map them. These regions of low mappability cause artificial 'valleys' of expression where the number of mapped reads drops close to zero. These valleys occur frequently in introns and lead to an underestimation of intron retention. To compensate for this we created a mappability index to correct these artificial valleys of expression and normalize intronic expression in low mappability regions. This index was calculated based on the Mus_musculus.GRCm38.74 reference genome. A sliding window of 40 bp and of step 10 bp was used to tile the reference genome. The genomic sequence in each window was extracted, a one base random mis-read was substituted and prepared in the format of sequencing reads (fastq). These artificially generated reads were then mapped against the reference

genome using the same parameters used for the input mRNA-seq data. Regions with resultant coverage at or poorer than 50% were considered to have poor mappability.

Reads were prepared by trimming adapters with a custom paired-end aware colour-space trimmer. Reads were mapped in single-end mode with CUSHAW3 allowing multi-mapping against a combined genome and junction transcriptome; the junction transcriptome was built with USeq MakeTranscriptome, allowing mapping across canonical and non-canonical combinations of known splice sites within genes. For each read pair, a unique correctly paired read was selected by custom code on best match measured by direction, distance separating reads, and mismatch count.

The depth of spliced read pairs was counted at both ends of the intron and the maximum taken; reads with at least 5 bp overhanging the splice site in both directions were considered.

The depth of intron cover was calculated from non-excluded bases within the intron. A trimmed mean of depth of cover was then calculated, including the centre 20% of values. All counts were performed with Bedtools. Coverage was assessed per-molecule where read pairs had no more than 120 bp separation.

Spearman correlation coefficients were calculated for each gene between intron retention values and their corresponding RNA-seq FPKM expression values across the 13 samples. For determining the effect of intron retention during the reprogramming stages described in Extended Data Fig. 6e, we performed Pearson Correlation analysis between intron retention values and their corresponding FPKM expression across a minimum of four samples. We also performed a similar correlation by randomizing gene expression values ten times to intron retention values for each indicated reprogramming stage to obtain the random level of correlation between intron retention and gene expression. This was used to calculate statistical significance.

Microarray data processing. Affymetrix HT Mouse Genome 430A microarray data from Polo *et al.*[6] (GEO accession number GSE42379) and Illumina MouseWG-6 v2.0 expression beadchip array data from Kojima *et al.*[20] (GEO accession number GSE46227) were analysed using R application and limma R package (v3.14.4). The probe intensity data were log transformed and quantile normalized and unannotated probes were removed.

miRNA sequencing and alignment. miRNA purification was performed according to the miRvana miRNA isolation kit (Ambion 1560) and quality validated using a Bioanalyser before sequence library preparation. Small RNA libraries were prepared for SOLiD next generation sequencing, with libraries sequenced to a depth of 27,420,558–118,946,232 tags (average 55,816,766 tags; up to 35 nucleotides in length), yielding a total of 725,617,952 tags (Supplementary Table 6). These tags were then mapped to the mouse genome (NCBI37/mm9 assembly) and miRNA-mapped tags determined as those overlapping with known miRNA loci (miRbase v18). Thus, using the tools and parameters detailed below, we were able to map 347,190,702 tags across the 13 libraries (47% of tags) (refer to 'Small RNA sequencing analysis pipeline' below for details).

Small RNA sequencing analysis pipeline. Identify and remove the adaptor sequence (maximum 25% mismatch with adaptor sequence). Retain tags with at least 20 nucleotides (nt) length, and at least 18 mean quality across the tag. Map tags to the mouse genome (mm9, NCBI37) and rRNA sequences using Bowtie (version 0.12.8) aligner: (i) command: bowtie -f -C -Q Sample,CV,qual–integers-quals -l 20 –nomaqround– maxbts 800 -y–chunkmbs 2048 -M -a –best –strata –snpfrac 0.01 –col-cqual –col-keepends –sam –mapq 20 –offrate 2 –threads 12 –shmem *ReferenceBowtieIndex.fa* Sample.csfasta; (ii) reference mouse genome (mm9 assembly), 18S rRNA (gi|374088232) and 28S rRNA (gi|120444900). (Refer to Supplementary Table 6 for tag counts.) Count the number of tags that overlap annotated miRNA as defined in miRNase version 18 (Tag length set between 20 and 26nt). For example: (i) miRBase annotates mature miRNA mmu-miR-XYZ on chromosome 1, starting at position 1,347 on the sense strand. All tags comply with the criteria below are assumed to mmu-miR-XYZ miRNA tags: 20–26 nt long; map to the sense strand of chromosome 1; start position between 1,344 and 1,350 inclusive (1,347 ± 3). Scale the number of tags assigned to each miRNA to correct for different library sizes. For example: (i) total tags mapped to miRNA loci in primary iPSC library = 9,618,934 and total tags mapped to miRNA loci in secondary iPSC library = 9,107,222, then a miRNA expression value is calculated as follows: if number of tags mapped to mmu-let-7a-5p miRNA in primary iPSC library is 36,868 then number of tags mapped to mmu-let-7a-5p miRNA in primary iPSC library after correcting for library size is (36,868/9,618,934) × 1,000,000 = 3,832.86; if number of tags mapped to mmu-let-7a-5p miRNA in secondary iPSC library is 47,890 then number of tags mapped to mmu-let-7a-5p miRNA in secondary iPSC library after correcting for library size is (47,890/9,107,222) × 1,000,000 = 5,258.46. Re-scale the library size using TMM method to compensate for sequencing real-estate effect[48]. Normalized counts for each miRNA are used for differential expression analysis.

Chromatin immunoprecipitation sequencing (ChIP-seq). *ChIP library generation.* ChIP was carried out as described previously[49]. 40–150 million cells were

fixed with 1% formaldehyde for 10 min at room temperature, scraped and stored as pellets ($-80\,^\circ$C). Samples were lysed at 20 million cells per ml in Farnham lysis buffer for 10 min followed by 10 million cells per ml in nuclear lysis buffer. The released chromatin was sheared to 100–500 bp (250 bp average) on ice using a SonicsVibraCell Sonicator equipped with a 3 mm probe. For each sample, 50 μl of solubilized chromatin was used as input DNA to normalize sequencing results and the remaining chromatin was immunoprecipitated with 10 μg of H3K4me3 (ab8580)[50], 10 μg H3K27me3 (Millipore 07-449)[51] or 10 μg H3K36me3 (ab9050)[51] antibodies, separately. Antibody–chromatin complexes were pulled down with 100 μl magnetic Protein G Dynal beads (Invitrogen) and washed six times. The chromatin was then eluted, reverse cross-linked at $65\,^\circ$C overnight and subjected to RNaseA/ proteinase K treatment. ChIP and input DNA was purified using a Qiagen Purification Column and quantified using a Quant-it dsDNA High Sensitivity Assay (Invitrogen).

High-throughput sequencing. Sequencing libraries were prepared according to Illumina ChIP-seq Library Preparation kit instructions. 50 ng of immunoprecipitated or input DNA was end-repaired, followed by the 3′ addition of a single adenosine nucleotide and ligation to universal library adapters. Ligated material was separated on a 2.0% agarose gel, followed by the excision of a 250–350 bp fragment and column purification using Qiagen gel purification kit. DNA libraries were prepared by PCR amplification (18 cycles). ChIP DNA libraries were sequenced using the Illumina HiSeq 2000 as per the manufacturer's instructions. Sequencing libraries was performed up to 2×101 cycles. Image analysis and base calling were performed with the standard Illumina pipeline version RTA 2.8.0.

Processing and alignment of ChIP-seq data to identify H3K4me3, H3K27me3 and H3K36me3 enriched peaks. ChIP-seq sequencing data were processed using the Illumina analysis pipeline and FastQ format reads were aligned to the NCBI37/ mm9 mouse reference using the Bowtie alignment algorithm[52]. Bowtie version 2.1.0 was used with the pre-set sensitive parameter to align ChIP sequencing reads from this study (refer to 'ChIP sequencing analysis pipeline' below for more details and Supplementary Table 6 for read counts) and ChIP-seq data set from Polo et al.[6] (GEO accession number GSE42477).

Peak calling algorithm. The MACS version 2.0.10 (model based analysis of ChIP-seq)[53] peak finding algorithm was used to identify regions of ChIP-Seq enrichment over background. Default parameters were used for H3K4me3, and broad peak parameters were used for H3K27me3 and H3K36me3 data (refer to 'ChIP sequencing analysis pipeline' below for details).

Peak annotation and processing. Multicov command from Bedtools v2.17.0 was used to obtain raw read counts within each histone mark peak identified by MACS and input reads within these peaks. The number of reads per kilobase of peak per million reads (RPKM) was calculated for each peak and the corresponding input levels of that peak. The RPKM values for the histone mark peak were then subtracted by those of the input RPKM values to obtain a final and background-adjusted RPKM value, as modified from ref. 54. Peak calls with background-adjusted RPKM values less than or equal to 0 were excluded from further analysis. The background-adjusted RPKM values were averaged across -2 kb to $+3$ kb of a gene TSS, as determined above, for downstream data analysis and visualization. Gene loci with an average background adjusted RPKM values less than 0.5 were considered negative for the presence of the histone mark. ngs.plot.r[55] software was used to generate read density heat maps and profiles. Read densities and enrichment scores per locus, where defined, were normalized to the total number of million uniquely mapped reads producing values in units of reads per million mapped reads (RPM).

Identification of differential histone mark changes associated with Fig. 3c–g. To determine a histone mark change during reprogramming, as shown in Fig. 3c–g, we first applied the following criteria for transcriptionally active and silent loci identification. Active locus: (i) H3K4me3$^+$H3K27me3$^-$H3K36me3$^{+/-}$ and gene expression values of $\log_2(\text{FPKM}) \geq 0.7226907$ for protein-coding genes or $\log_2(\text{FPKM}) \geq -1.515307$ for lncRNAs, as determined in Extended Data Figs 8a and 9a, respectively. Silent locus: (i) H3K4me3$^{+/-}$H3K27me3$^+$H3K36me3$^-$ and gene expression values of $\log_2(\text{FPKM}) < 0.7226907$ for protein-coding genes or $\log_2(\text{FPKM}) < -1.515307$ for lncRNAs; (ii) H3K4me3$^-$H3K27me3$^-$H3K36me3$^-$ (that is, no mark) and $\log_2(\text{FPKM}) < 0.7226907$ for protein-coding genes or $\log_2(\text{FPKM}) < -1.515307$ for lncRNAs.

Only histone marks that follow the criteria described above were considered for further analysis.

We next grouped samples as follows. Group 1, secondary MEF; group 2, D2H; group 3, D5H–D11H; group 4, D16H–D18H, D16L; group 5, D21L and D21Ø; group 6, secondary iPSC, ESC and primary iPSC.

We then only examined histone mark modifications where a change was observed from secondary MEF to a minimum of two samples from within groups 3, 4 or 6. In cases where a gene switched transcriptional activity, for example, changing from active to silent or vice versa, our analysis only focused on genes showing stage-specific expression by RNA-seq.

ChIP sequencing analysis pipeline. Trim sequence (filter out 3′ adaptor, and remove last 2 bases and 3 extra bases if it matches with adaptor sequence). Mapping sequences to mouse genome (mm9/NCBI37) using Bowtie: (i) command: bowtie2 -p 8 –sensitive -x mm9/mm9 -1 sequence.reads_R1.fastq -2 sequence.reads _R2.fastq –S sample.sam. (Refer to Supplementary Table 6 for read counts.) Peak calling algorithm MACS: (i) command for H3K4me3: macs2 callpeak -t chromatin. mark.file.bam -c input.sample.file.bam -f BAMPE -g mm -n [directory] –nomodel –shiftsize 73 -B; (ii) command for H3K27me3 and H3K36me3: macs2 callpeak –t chromatin.mark.file.bam –c input.sample.file.bam –broad –f BAMPE -g mm -n [directory] –nomodel –shiftsize 73 -B. (Refer to Supplementary Table 6 for read counts.) Normalize unique mapped read values to library size. Annotate peaks to mouse genome (mm9/NCBI37).

DNA methylation analysis. *MethylC-seq library generation.* Five micrograms of genomic DNA was mixed with unmethylated cl857 Sam7 Lambda DNA (Promega, Madison, WI, USA). The DNA was fragmented by sonication to 300–500 bp with a Covaris S2 system (Covaris) followed by end repair with the End-It DNA End-Repair kit (Epicentre). Paired-end universal library adaptors provided by Illumina (Illumina) were ligated to the sonicated DNA as per manufacturer's instructions for genomic DNA library construction. Ligated products were purified with AMPure XP beads (Beckman, Brea, CA). Adaptor-ligated DNA was bisulphite treated using the EpiTect Bisulphite kit (Qiagen) following the manufacturer's instructions and then PCR amplified using PfuTurboCx Hotstart DNA polymerase (Agilent, Santa Clara, CA) with the following PCR conditions (2 min at $95\,^\circ$C, 4 cycles of 15 s at $98\,^\circ$C, 30 s at $60\,^\circ$C, 4 min at $72\,^\circ$C then 10 min at $72\,^\circ$C). The reaction products were purified using the MinElute gel purification kit (Qiagen). The sodium bisulphite non-conversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the lambda genome.

High-throughput sequencing. MethylC-seq DNA libraries were sequenced using the Illumina HiSeq 2000 as per the manufacturer's instructions. Sequencing was performed up to 2×101 cycles. Image analysis and base calling were performed with the standard Illumina pipeline version RTA 2.8.0.

Processing and alignment of MethylC-seq data to identify methylated cytosines. MethylC-seq sequencing data was processed using the Illumina analysis pipeline and FastQ format reads were aligned to the NCBI37/mm9 mouse reference using the Bismark/Bowtie alignment algorithm[52,56]. Paired-read MethylC-seq sequences produced by the Illumina pipeline in FastQ format were trimmed with trim threshold 1,500, which removed the last two bases from sequences that were not trimmed, and removed three bases from sequences that were trimmed. The Bismark package version 0.7.7 was used as the aligner (refer to 'Methylome sequencing analysis pipeline' below for more details).

Since up to six independent libraries from each biological replicate were sequenced, we first removed duplicate reads. Subsequently, the reads from all libraries of a particular sample were combined. Unique read alignments were then subjected to post-processing. The number of calls for each base at every reference sequence position and on each strand was calculated. All results of aligning a read to both the Watson and Crick converted genome sequences were combined. The CpG methylation levels were calculated using bisulphite conversion rates by (number of not converted Cs/read depth) for each position.

Identification of methylated cytosines. At each reference cytosine the binomial distribution was used to identify whether at least a subset of the genomes within the sample were methylated, using a 0.01 FDR corrected P value. We identified methyl cytosines while keeping the number of false-positive methylcytosine calls below 1% of the total number of methyl cytosines we identified. The probability P in the binomial distribution B (n,P) was estimated from the number of cytosine bases sequenced in reference cytosine positions in the unmethylated Lambda genome (referred to as the error rate: non-conversion plus sequencing error frequency). We interrogated the sequenced bases at each reference cytosine position one at a time, where read depth refers to the number of reads covering that position. For each position, the number of trials (n) in the binomial distribution was the read depth. For each possible value of n we calculated the number of cytosines sequenced (k) at which the probability of sequencing k cytosines out of n trials with an error rate of P was less than the value M, where $M\times$ (number of unmethylated cytosines) $<0.01\times$ (number of methylated cytosines) and if the error rate of P was over 0.01, we assumed the cytosine was not methylated. In this way, we established the minimum threshold number of cytosines sequenced at each reference cytosine position at which the position could be called as methylated, so that out of all methyl cytosines identified no more than 1% would be due to the error rate.

Identification of differentially methylated regions (DMRs). DMRs were identified using a sliding window approach of 300 bp, sliding every 30 bp. Windows showing differences above 45% between any sample and a minimum of 5 CpGs were considered differentially methylated. 13,1540 differentially methylated windows were identified. Differentially methylated windows were merged to obtain an average methylation level or differential methylation value, relative to secondary MEFs,

per annotated gene locus. Analysis was confined to −1 kb to +1 kb region of TSS as we found this to be the region frequently spanning hypomethylation for key ESC genes.

Methylome sequencing analysis pipeline. Trim sequence (filter out 3′ adaptor, and remove last two bases and three extra bases if it matches with adaptor sequence). Mapping sequences to mouse genome (mm9/NCBI37) using Bismark/Bowtie using the following parameters: (i) command: -e 90 -n 2 -l 32 -X 550; (ii) sequence reads are first transformed into fully bisulphite-converted forward (C>T) and reverse read (G>A conversion of the forward strand) versions. They are then aligned to similarly converted versions of the genome (also C>T and G>A converted). Sequence reads that produce a unique best alignment from the four alignment processes against the bisulfite genomes (which are running in parallel) are then compared to the normal genomic sequence and the methylation state of all cytosine positions in the read is inferred. A read is considered to align uniquely if one alignment exists that has fewer mismatches to the genome than any other alignment (or if there is no other alignment). (Refer to Supplementary Table 6 for read counts and methylated cytosine distribution.) Remove duplicates. Calculate base-by-base methylation level and final CpG methylation counts. (Refer to Supplementary Table 6 for methylated CpG counts.) Integrate CpG methylation level of positive and negative strand. Adjust methylation level using bisulphite conversion rate from unmethylated Lambda control.

Global proteomics. *Sample preparation for MS analysis.* Cells were harvested by centrifugation and lysed in 8 M urea (100 mM triethyl ammonium bicarbonate, pH 8.2, with protease and phosphatase inhibitors). Proteins (~1 mg) were first reduced/alkylated and digested for 4 h with Lys-C. The mixture was then diluted fourfold to 2 M urea and digested overnight with sequencing grade trypsin (Promega) in substrate/enzyme ratio of 50:1 (w/w). Digestion was quenched by acidification with formic acid (final concentration 10%). Resulting peptides were subsequently desalted by solid phase extraction (Sep-pack Vac C18 cartridges, Waters), dried down and then re-suspended in TEAB buffer 100 mM to a final concentration of ~1 mg ml^{-1}. An aliquot of 100 μg of each sample was chemically labelled with Tandem Mass Tag (TMT) reagents (Thermo Fisher) according to the manufacturer's instructions. Data for all samples were normalized to an internal standard (ISTD) made up of equal proportions of the samples (refer to 'Global proteome analysis pipeline' below for details). Before the mass spectrometric analysis, both the TMT-labelled peptides mixtures were fractionated as described elsewhere[57]. The SCX system consisted of an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn, Germany) with one C18 Opti-Lynx (Optimized Technologies, OR) trapping cartridges and a Zorbax BioSCX-Series II column (0.8 mm inner diameter 50 mm length, 3.5 mm). The labelled peptides were dissolved in 10% formic acid and loaded onto the trap columns at 100 μl min^{-1} and subsequently eluted onto the SCX column with 80% acetonitrile (ACN; Biosolve, The Netherlands) and 0.05% formic acid. A total of 50 SCX fractions (1 min each; that is, 40 μl elution volume) were collected and used for subsequent LC-MS/MS analysis.

Mass spectrometric analysis. We performed nanoflow LC-MS/MS using an LTQ-Orbitrap Velos mass spectrometer (Thermo Electron, Bremen, Germany) coupled to an Agilent 1200 HPLC system (Agilent Technologies). SCX fractions were dried, reconstituted in 10% FA and delivered to a trap column (ReproSil C18, (Dr Maisch GmbH, Ammerbuch, Germany); 20mm × 100μm inner diameter, packed in-house) at 5 μl min^{-1} in 100% solvent A (0.1 M acetic acid in water). Next, peptides eluted from the trap column onto an analytical column (ReproSil-Pur C18-AQ (Dr Maisch GmbH, Ammerbuch, Germany); 40 cm length, 50 μm inner diameter, packed in-house) at approximately 100 nl min^{-1} in a 90 min or 3 h gradient from 0 to 40% solvent B (0.1 M acetic acid in 8:2 (v/v) ACN/water). The eluent was sprayed via distal coated emitter tips butt-connected to the analytical column. The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra (from *m/z* 350 to 1,500) were acquired in the Orbitrap with a resolution of 30,000 FHMW at 400 *m/z* after accumulation to target value of 500,000 in the linear ion trap (maximum injection time was 250 ms). After the survey scans, the ten most intense precursor ions at a threshold above 5,000 were selected for MS/MS with an isolation width of 1.2 Da after accumulation to a target value of 30,000 (maximum injection time was 50 ms). Peptide fragmentation was carried out by using higher-energy collisional dissociation (HCD) with an activation time of 0.1 ms and a normalized collision energy of 45%. Fragment ions analysis was performed in the Orbitrap with a resolution of 7,500 FHMW and a low mass cut-off setting of 100 *m/z*.

Data processing. MS raw data were processed with Proteome Discoverer (version 1.3, Thermo Electron). Peptide identification was performed with Mascot 2.3 (Matrix Science) against a concatenated forward-decoy UniPROT database supplemented with all the frequently observed contaminants in MS (version 5.62). The following parameters were used: 50 ppm precursor mass tolerance, 0.02 Da fragment ion tolerance, up to two missed cleavages, carbamidomethyl cysteine as fixed modification, and oxidized methionine and TMT modification on N-Term and lysine

as variable modifications. Finally, we performed a deconvolution of the high-resolution MS2 spectra, by which all the fragment ions isotopic distributions were converted to an *m/z* value corresponding to the monoisotopic single charge. Reporter-ion-based quantification method was chosen in Proteome Discoverer, with the following requirements for reporter ion integration in the MS2 spectra: mass accuracy of maximum 20 ppm, peptide ratio maximum limit 100. To minimize ratio distortion due to the presence of more than one peptide species within the precursor ion isolation width, we also rejected the quantification of MS/MS spectra having a co-isolation higher than 30%. Finally, results were filtered using the following criteria: (i) mass deviations of ±5 ppm; (ii) mascot ion score of at least 25; (iii) a minimum of 7 amino acid residues per peptide; and (iv) position rank 1 in Mascot search. As a result, we obtained peptide FDRs of 0.3% for the mix 1 and 0.5% for the mix 2, which correspond to a protein FDR of 1% for the overlapping protein identification of the two 6-plex analyses. Finally, peptide ratios were log$_2$ transformed and normalized by median subtraction (refer to 'Global proteome analysis pipeline' below for details).

Global Proteome analysis pipeline. Sample mix composition: (i) ISTD: mixture in 1:1 ratio of secondary MEF, D2H, D5H, D8H, D11H, D16H, D18H, secondary iPSC, ESC, primary iPSC; (ii) mix1: secondary MEF, D2H, D5H, D8H, D11H and ISTD; (iii) mix2: D16H, D18H, secondary iPSC, ESC, primary iPSC and ISTD; (iv) mix3: D16L, D21L, D21ø and ISTD. Raw data processing (for example, noise filtering, de-isotoping, deconvolution) by using Proteome Discoverer 1.3 Software (Thermo): (i) mix1: 731,645 MS2 events; (ii) mix2: 725,642 MS2 events; (iii) mix3: 908,982 MS2 events. Mapping MS2 spectra to peptide sequences by using Mascot 2.3 search engine (Matrix Science), and the following parameters were used for database search: (i) mass tolerance of 50 ppm and 0.02 Da for precursor; (ii) up to two missed cleavages; (iii) cysteine carbamidomethylation as fixed modification; (iv) methiodine oxidation, TMT modification on lysine and peptide N termini as variable modifications; (v) concatenated forward-decoy database supplemented with all the frequently observed contaminants in MS (Uniprot v_2011_07_*Mus musculus*) was used. Filtered identification at a false discovery rate (FDR) lower than 1%. Peptide spectrum matches (PSMs) are filtered based on the following criteria in order to obtain an FDR <1%: (i) peptide length >6 amino acids; (ii) peptide rank = 1; (iii) ion score >25; (iv) delta mass <5 ppm; (v) the filtered identifications are summarized as follows: mix1, 199,373 PSMs, 39,518 peptides, 5,943 proteins; mix2, 220,729 PSMs, 46,206 peptides, 6,408 proteins; mix3, 153,869 PSMs, 40,838 peptides, 6,136 proteins. Reporter ion-based quantification is performed by using Proteome Discoverer 1.3 Software (Thermo): (i) relative quantification is performed dividing the MS2 intensities of the reporter ions of a given sample by the internal standard mixture (sample x/ISTD). Protein ratios are then calculated as the median of the peptides ratios within the same protein, and peptide quantification is accepted only if the reporter ions mass deviation is <20 ppm; the peptide is labelled both at the N-terminal and at the lysine residues (when present); the precursor ion shows a co-isolation interference <30%. (Refer to Supplementary Table 6 for number of quantified proteins that passed the filters).

Cell surface proteomics. *Sample preparation for MS analysis.* A simplified version of the cell surface capture (CSC) protocol introduced by Wollscheid *et al.*[58] was applied to identify *N*-glycosylated surface proteins over the project time course. Fixed quantities of protein (5 mg), as determined by a duplicate DC protein assay (Bio-Rad), were used in place of cell counts to determine the volumes of cell lysate to process further.

Mass spectrometric analysis. Vacuum centrifugation was performed on a volume of glycopeptide mixture calculated to be derived from 2 mg of total protein. After the volume was concentrated to several microlitres, it was then adjusted to 11 μl with 0.1% formic acid and transferred to a well of a 96-well plate, which was subsequently placed in an EASY-nLC nano LC pump (Proxeon) connected to a microcolumn. Microcolumns were created from sections of capillary-scale nanoflow 75 μm internal diameter fused silica tubing (Polymicro Technologies) pulled to a fine tip using a P-2000 laser puller (Sutter Instruments). They each were packed to a length of 10 cm with 5 μm Luna C18 resin (Phenomenix) using a pressure vessel, then flushed for 15 min with methanol. Microcolumns were regenerated with buffer A (5% acetonitrile and 0.1% formic acid in HPLC-grade water from Fisher) before loading of sample by the nano LC pump. Each chromatography session began with a linear gradient elution of 5% to 25% buffer B (95% acetonitrile and 0.1% formic acid in HPLC-grade water from Fisher) over 45 min followed by a linear gradient of 25% to 80% buffer B over 9 min. A flow rate of 300 nl min^{-1} was maintained. Peptides were analysed using nanospray ionization on an Orbitrap-Velos mass spectrometer (Thermo). MS1 and MS2 spectra were acquired with the instrument operating in the data-dependent mode of one MS scan (on the Orbitrap) followed by up to ten MS2 scans (on the LTQ-Velos) when triggered by ion signals above a threshold of 500. Fragmentation was accomplished using collision-induced association. Three LC-MS replicates were performed for each of the selected time points.

Database searching and analysis. All MS2 spectra were searched using the SEQUEST algorithm and the International Protein Index (IPI) mouse database (Version 3.84) with the reversed protein sequences appended as decoys. Confidences in identifications of peptides (of at least seven amino acids in length) were evaluated using the Statquest probabilistic model[59] and further filtered to within a mass tolerance of 20 ppm using the accurate ion masses generated by the Orbitrap, thereby attaining an estimated false positive rate of 2%. Any identified peptides were then excluded if they did not include the *N*-glycosylation consensus sequon NXS/T or did not demonstrate the asparagine to aspartic acid deamidation of 0.986 Da resulting from the treatment with PNGaseF. Relative quantities of cell surface proteins were assessed by spectral counting or through use of matching global proteomic quantitative data where possible (refer to 'Cell surface proteome analysis pipeline' below for details).

Cell surface proteome analysis pipeline. Samples and controls: (i) samples: secondary MEF, D2H, D5H, D8H, D11H, D16H, D18H, secondary iPSC; (ii) controls: ESC, primary iPSC; (iii) three replicates per sample. Raw data processing (charge state assignment) using 'extractms' v.2 (rev.11): (i) replicate set 1: 219,736 MS2 events; (ii) replicate set 2: 219,467 MS2 events; (iii) replicate set 3: 238,273 MS2 events. Mapping MS2 spectra to peptide sequences by using Sequest v.27 (rev.9) and in-house supporting programs search engine (Matrix Science), and the following tolerances and parameters were used for database search: (i) peptide mass tolerance of 3.0 Da; (ii) up to one missed cleavage; (iii) cysteine carbamidomethylation as fixed modification; (iv) asparagine deamidation as a variable modification; (v) international protein index (IPI) mouse database (Version 3.84), with appended reversed (decoy) database was used. Filtered identification at a false discovery rate (FDR) lower than 2%. Peptide identifications were filtered as follows to obtain an FDR score <2%: (i) initial confidence estimation using STATQUEST methodology; (ii) precursor delta mass <20 ppm; (iii) peptide sequences contain *N*-glycosylation 'sequon' (NXS or NXT, where X is any amino acid except proline); (iv) the filtered identifications are summarized as follows: 14,917 spectral counts; 896 identified glycopeptides; 432 identified glycoproteins. A total of 432 cell surface proteins passed the filters. A total of 185 overlapped with the global proteomics quantitative data set.

43. Behringer, R. R., Gertsenstein, M., Nagy-Vintersten, K. & Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2013).
44. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
45. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007).
46. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
47. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
48. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
49. O'Geen, H., Echipare, L. & Farnham, P. J. in *Epigenetics Protocols* **791**, 265–286 (Humana, 2011).
50. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863–868 (2009).
51. Wang, T., Liu, J., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature Protocols* **7**, 1728–1740 (2012).
52. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **12**, R22 (2011).
53. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature Protocols* **7**, 1728–1740 (2012).
54. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
55. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
56. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
57. Gauci, S. *et al.* Lys-N and trypsin cover complementary parts of the phospho-proteome in a refined SCX-based approach. *Anal. Chem.* **81**, 4493–4501 (2009).
58. Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnol.* **27**, 378–386 (2009).
59. Kislinger, T. *et al.* PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol. Cell. Proteomics* **2**, 96–106 (2003).
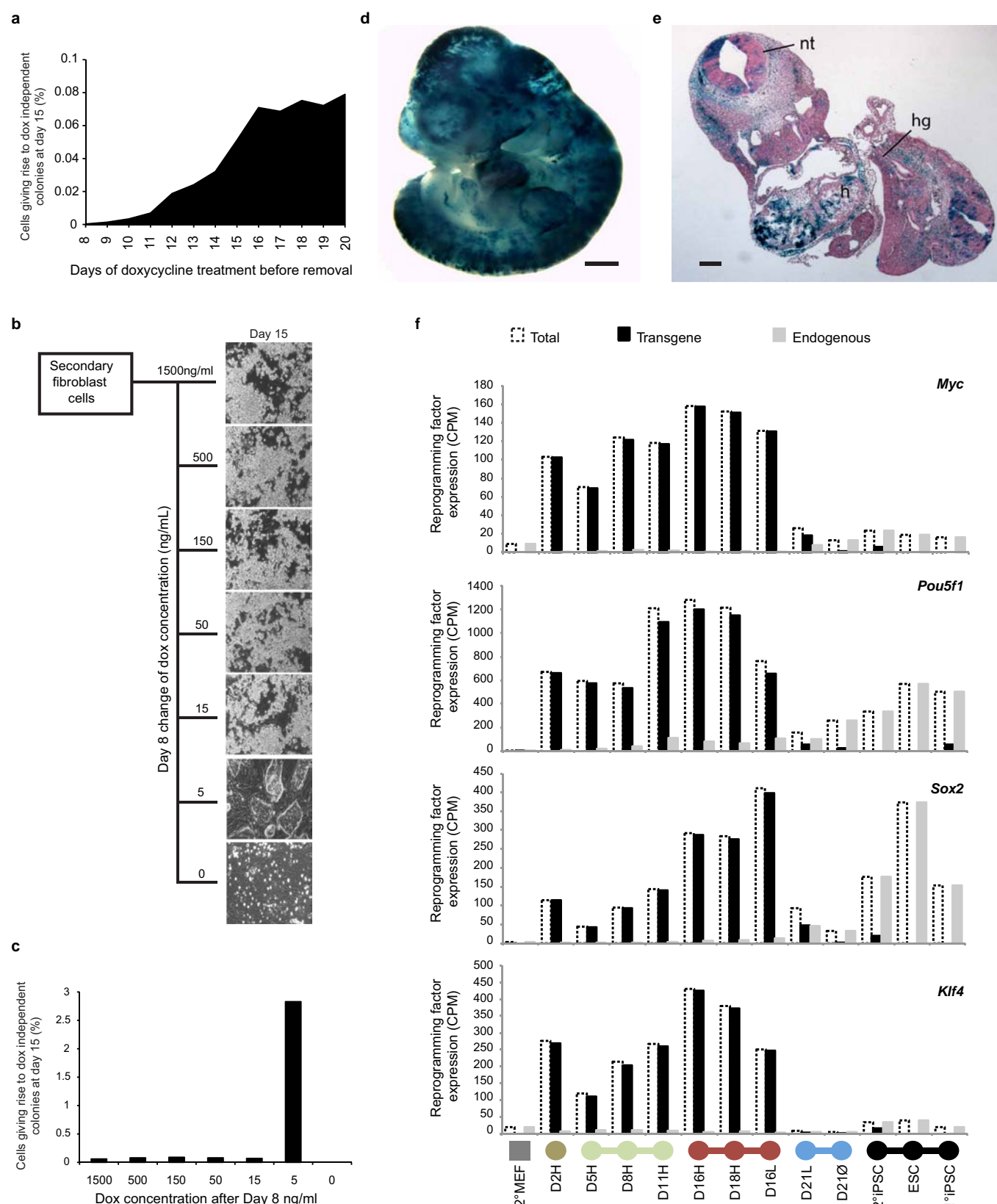
Extended Data Figure 1 | Effects of lowering doxycycline on reprogramming cells. a, Frequency of doxycycline-independent pluripotent cells obtained when 1B secondary MEFs were reprogrammed in 1,500 ng ml^{-1} doxycycline until the indicated day. **b**, Morphology of cells at day 15 after lowering the doxycycline concentration from 1,500 ng ml^{-1} to levels as indicated on day 8 of reprogramming. **c**, Clonal efficiency measurement at day 15 of reprogramming after lowering the doxycycline concentration on day 8 to the level indicated. **d, e**, 1B secondary iPSCs show widespread contribution to all germ layers of chimaeric embryos. Whole-mount view (**d**) and transverse section of E10.5 diploid chimaera (**e**). Embryo is representative of $n = 6$ chimaeric embryos with strong (>75%) iPSC donor cell contribution. h, heart; hg, hindgut; nt, neural tube. Scale bars, 750 μm (**d**) and 400 μm (**e**). **f**, RNA-seq analysis of transgene and endogenous expression levels during reprogramming. CPM, counts per million.
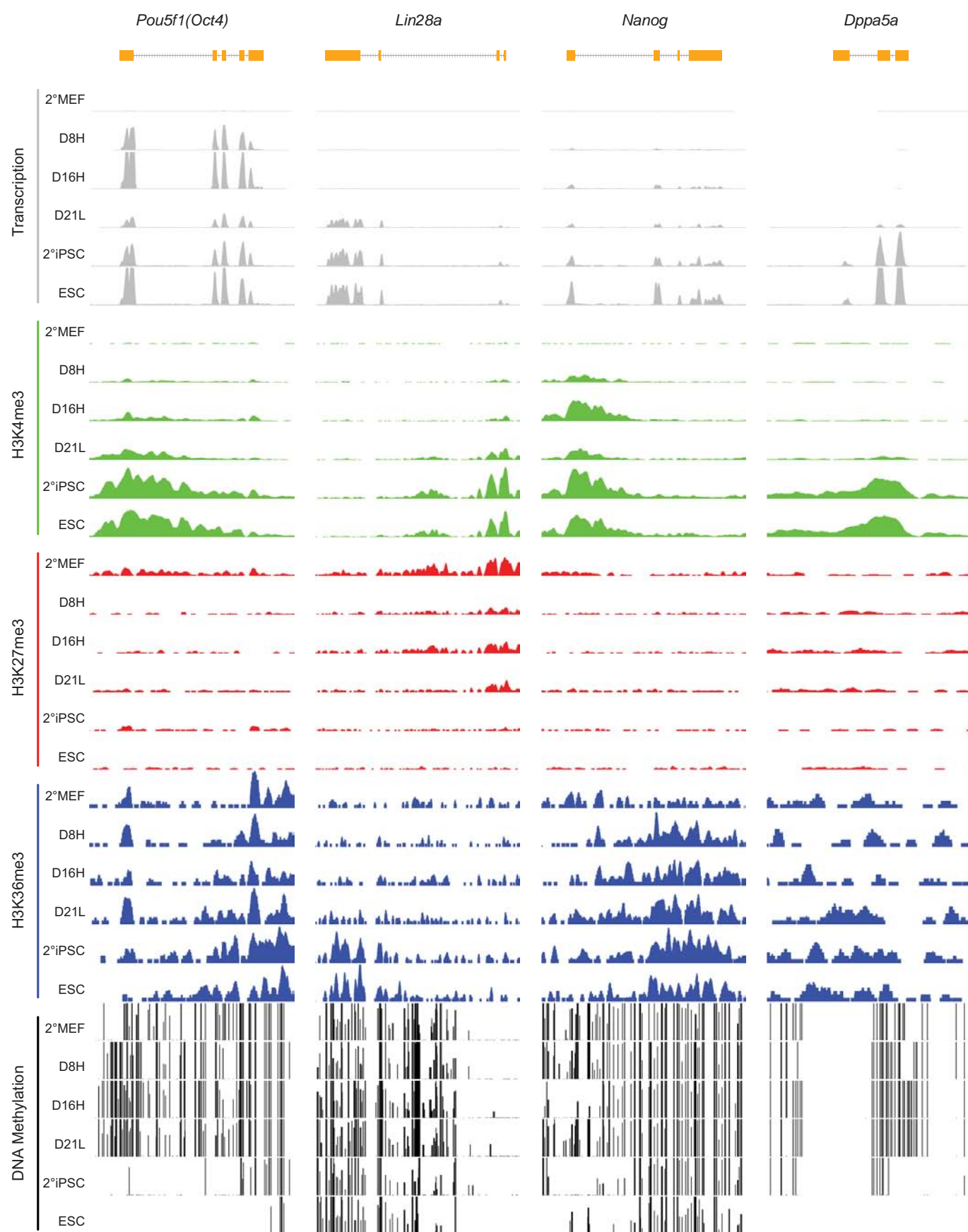
Extended Data Figure 2 | Locus-specific sequencing data. Read coverage histograms representing gene expression and epigenetic status at the genomic loci of selected ESC-associated genes.

a, Long RNA-Seq (mRNA and lncRNA): Hierarchical Clustering

c, Global CpG Methylome

b, Small RNA (miRNA)

Long RNA (protein coding)

Cell Surface Proteome

d, H3K4me3

H3K36me3

H3K27me3
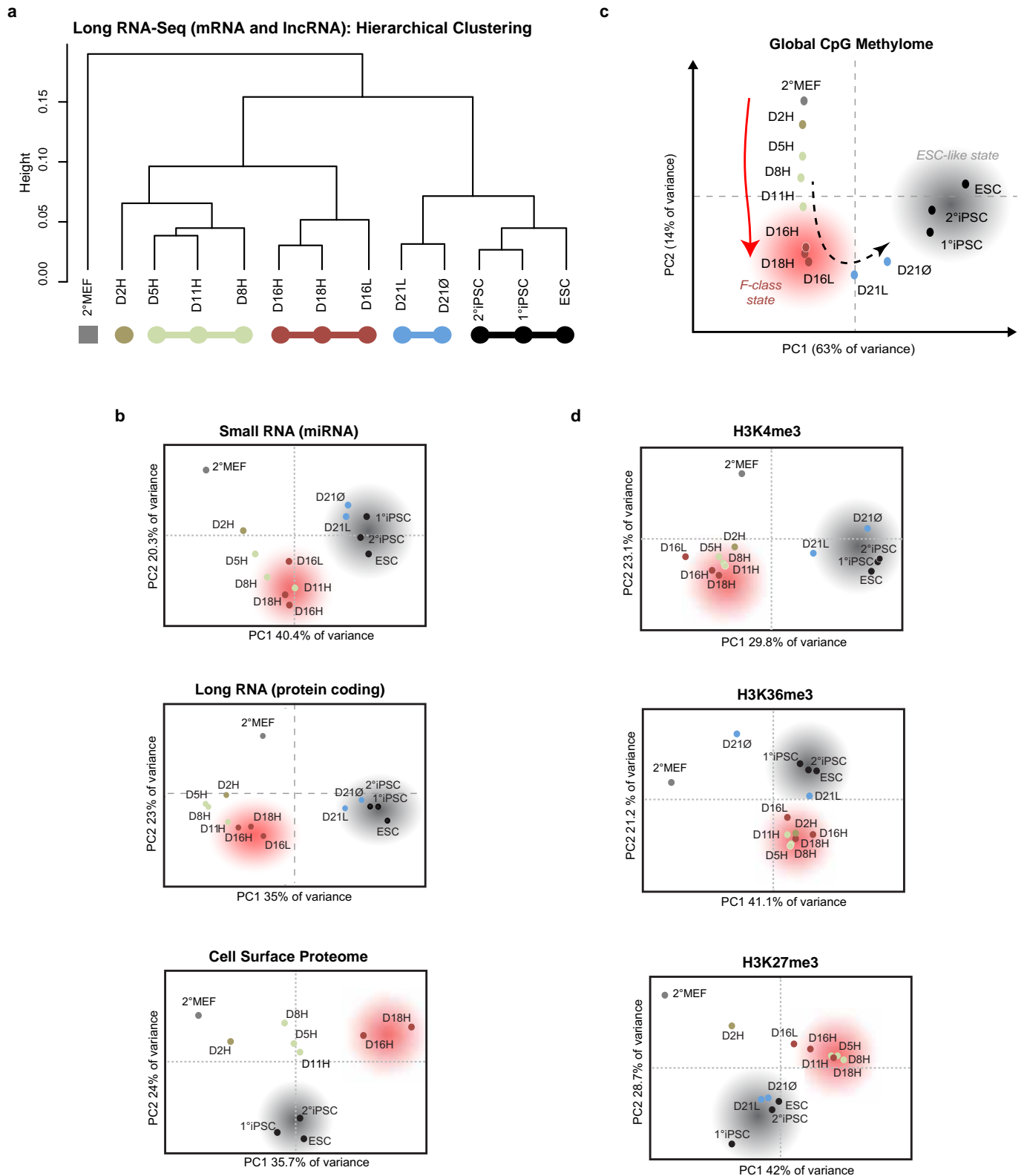
Extended Data Figure 3 | Hierarchical clustering and principal component analysis (PCA) for multi-omics analyses. a, Pearson correlation complete linkage hierarchical clustering of long RNA-seq data set. Colour coding indicates the grouping of samples based on clustering. **b–d,** PCA performed on each platform (10 neighbours for *k*-value nearest neighbour (KNN) imputation). Short RNA-seq platform PCA was performed on miRNAs (**b**). Long RNA-seq platform PCA was performed on protein-coding transcripts (**b**). Cell surface proteome PCA represents proteins detected by cell surface focused mass spectrometry analysis (**b**). **c,** PCA of global CpG methylation analysis. Red arrow follows the high-doxycycline sample trajectory; black dashed arrow follows D8H through low-doxycycline trajectory. Low-doxycycline samples D21L and D21∅ are highlighted in blue to indicate that compared to other platforms they do not project with ESC/iPSC (see text for further details). **d,** H3K4me3, H3K36me3 and H3K27me3 PCAs represent genome-wide enriched regions at annotated genes.
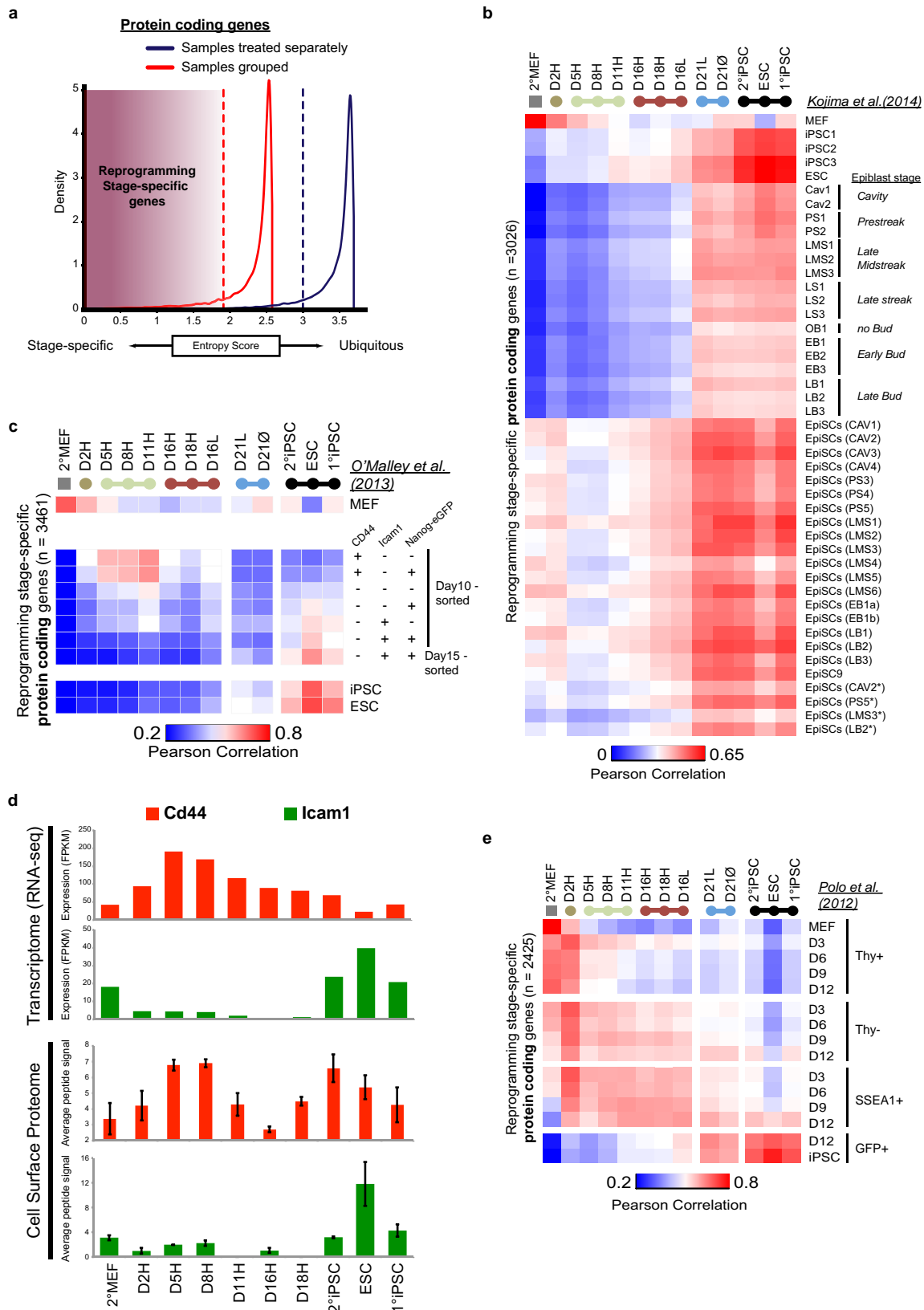
Extended Data Figure 4 | Integration of gene expression data from 1B reprogramming and other transcriptome data sets. **a**, Distribution of the entropy score of protein-coding gene expression for individual samples (blue) and sample groups (red) indicated as probability density curve. **b**, Pearson correlation analysis of 1B secondary reprogramming sample protein-coding gene expression with transcriptomes of early embryonic stages and epiblast stem cells (EpiSCs) derived from a range of developmental stages[20]. **c**, Pearson correlation analysis of 1B secondary reprogramming sample protein-coding gene expression with transcriptome of sorted secondary reprogramming intermediates[8]. **d**, Expression of CD44 and Icam1 markers during 1B reprogramming. Error bars represent standard error of the mean. **e**, Pearson correlation analysis of 1B reprogramming sample protein-coding gene expression with sorted reprogramming and pluripotent cells from the Col1a1 primary reprogramming system[6].

a

b

c

d
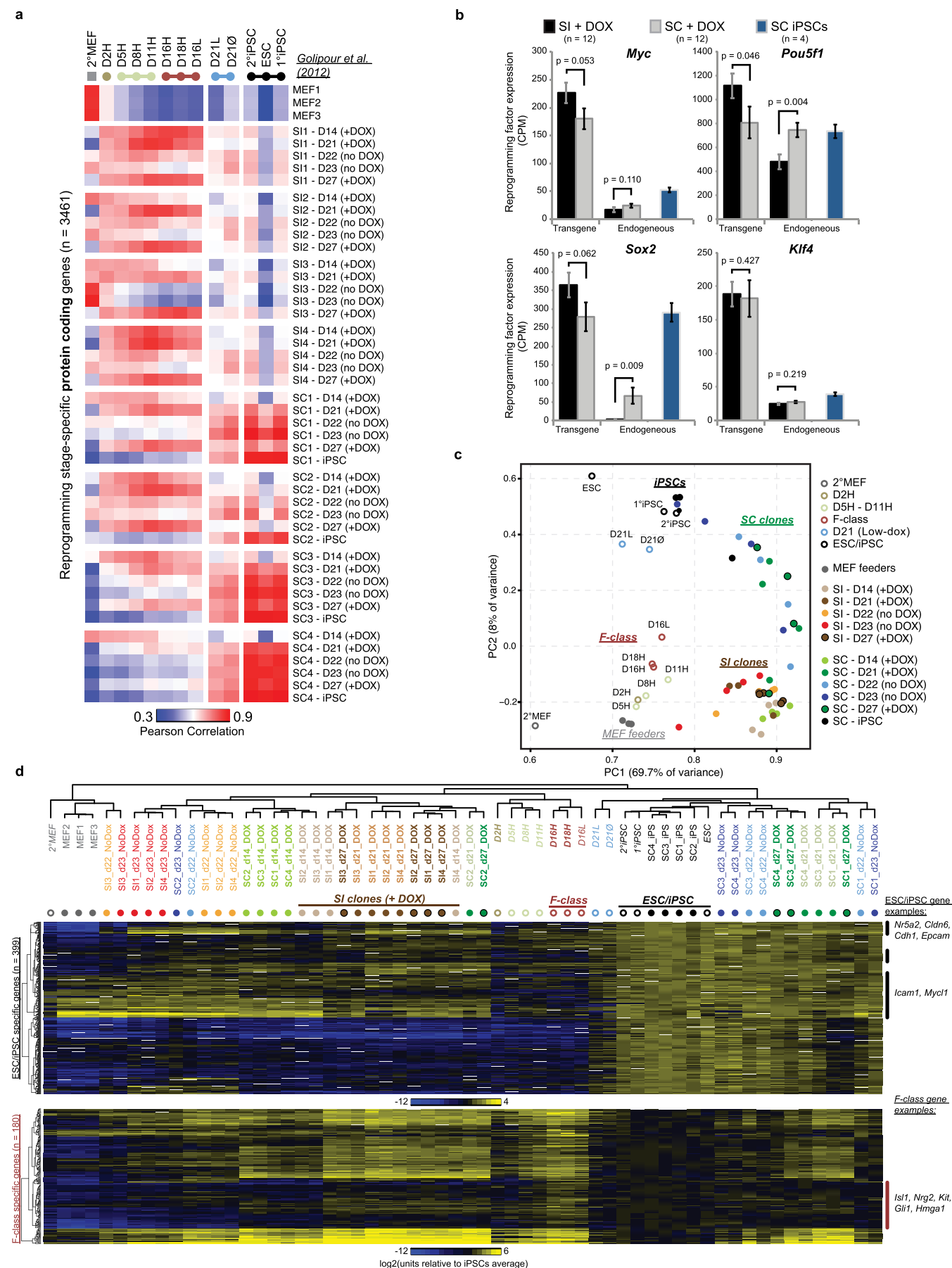
Extended Data Figure 5 | Effect of *Oct4, Sox2, Klf4* and *Myc* expression level on reprogramming outcomes. a, Pearson correlation analysis of RNA-seq data from 1B reprogramming samples and reprogramming clones from ref. 7 that are competent or incompetent to become factor-independent secondary iPSC (SC and SI clones, respectively). **b**, Transgene and endogenous gene expression determined by RNA-seq for *Myc*, *Pou5f1* (*Oct4*), *Sox2* and *Klf4* in SC and SI clones[7]. Bar graphs represent average expression of doxycycline-treated samples or SC iPSCs. Error bars represent standard error of the mean. Student's *t*-test was used for statistics. **c**, PCA of protein-coding stage-specific genes from Fig. 2c, comparing 1B reprogramming samples and secondary reprogramming clones from ref. 7. F-class cells cluster separately from SI and SC clones. Moreover, 1B reprogramming follows a different trajectory than SI and SC clones towards iPSCs. Colour coding indicates the grouping of samples. **d**, Pearson correlation complete linkage hierarchical clustering of 1B reprogramming samples and SI and SC secondary reprogramming clones. Clustering was performed on protein-coding stage-specific genes and based on FPKM values normalized to the averaged ESC/iPSCs values from the respective study. Heat maps show stage-specific protein-coding gene expression belonging to iPSC/ESC (top heat map) and F-class (bottom heat map) genes. Clusters and genes on the right of each heat map highlight genes that show a different expression pattern between F-class and doxycycline-treated SI clones. For gene lists associated with **d**, refer to Supplementary Table 1.
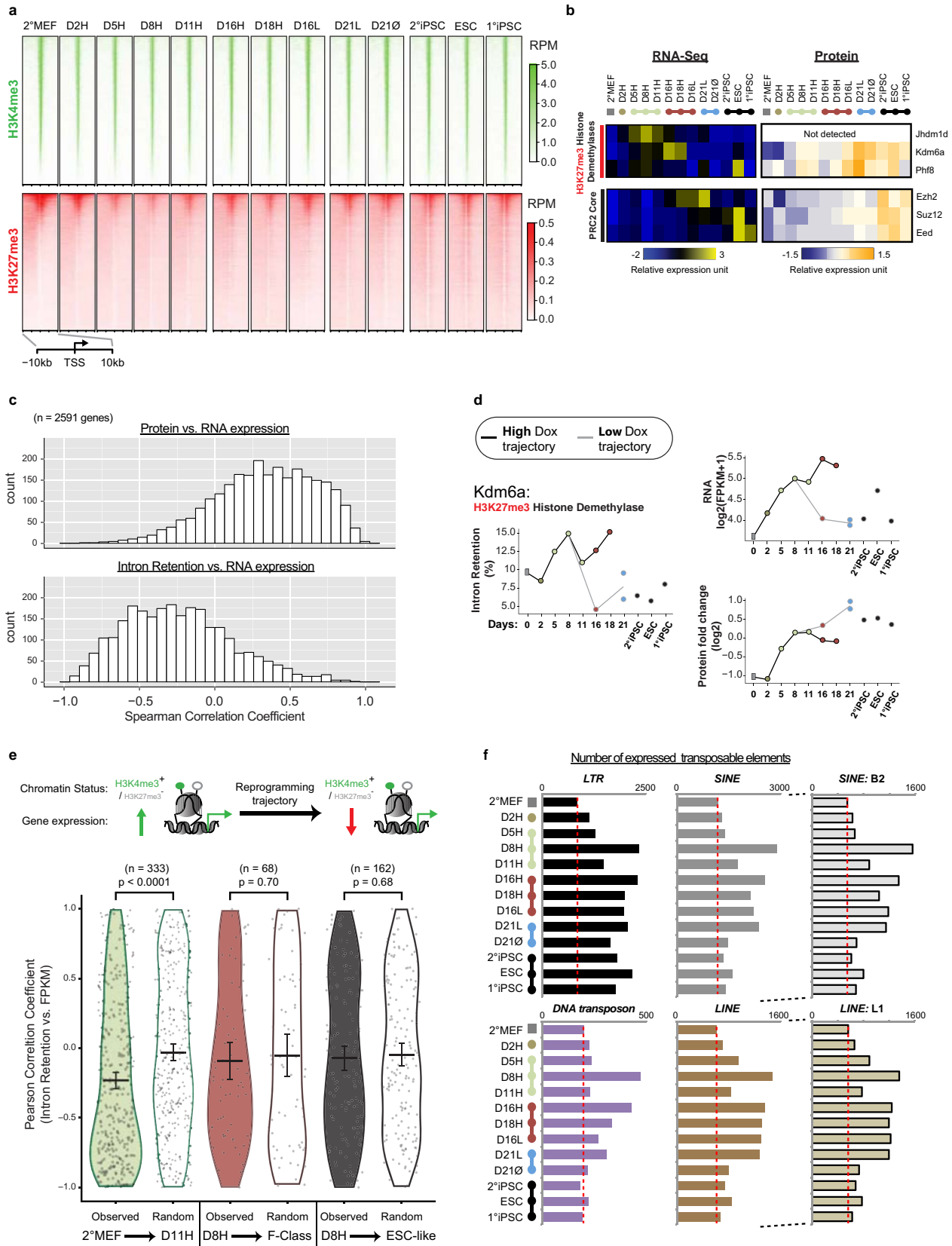
Extended Data Figure 6 | Global analysis of histone mark and intron retention changes during reprogramming. a, Intensity plots of genes associated with H3K4me3 (green) and H3K27me3 (red) ±10 kb of annotated TSSs. **b**, Heat map representation of PRC2 components and histone demethylase expression at the RNA (RNA-seq) and protein level. **c**, Correlation of gene transcription with protein and intron retention for genes that exhibit intron retention from Fig. 2c. **d**, Correlation of intron retention, RNA expression and protein level for Kdm6a. **e**, Violin plots comparing observed and random Pearson correlations of intron retention versus gene FPKM at reprogramming stages. Bars represent average Pearson correlation coefficients. Error bars represent standard error of the mean. Student's *t*-test was used for statistics. **f**, Number of expressed transposable elements during reprogramming.

Extended Data Figure 7 | Tracking secondary MEF histone mark changes during reprogramming from one sample to another. a, Pie-chart diagram tracking the histone mark changes using secondary MEF and secondary iPSCs as reference points. Each histone mark is colour coded: H3K4me3, green; H3K4me3H3K27me3, orange; H3K27me3, red; no mark, grey. Loci were tracked from their start (2°MEF) and end (2°iPSCs) histone signatures. **b–g**, Tracking bar graphs of histone mark changes. The histone mark change is shown at the top of each set of 12 histograms. Bars represent number of genes whose mark changed for the time point indicated at the top of the

individual histogram, and which of these genes carry the same mark at the other time points (x axis). For example, in **b** '2°MEF (H3K4me3/ H3K27me3→H3K4me3)' the histogram shows the number of genes that were bivalent in secondary MEFs but changed to H3K4me3 monovalent at another time point. In the case of the small histogram labelled D2H, the black-framed green bar represents the number of loci that showed this change from secondary MEFs at D2H and the bars for all the other samples indicate how many of these D2H loci were also H3K4me3[+] in the other samples.
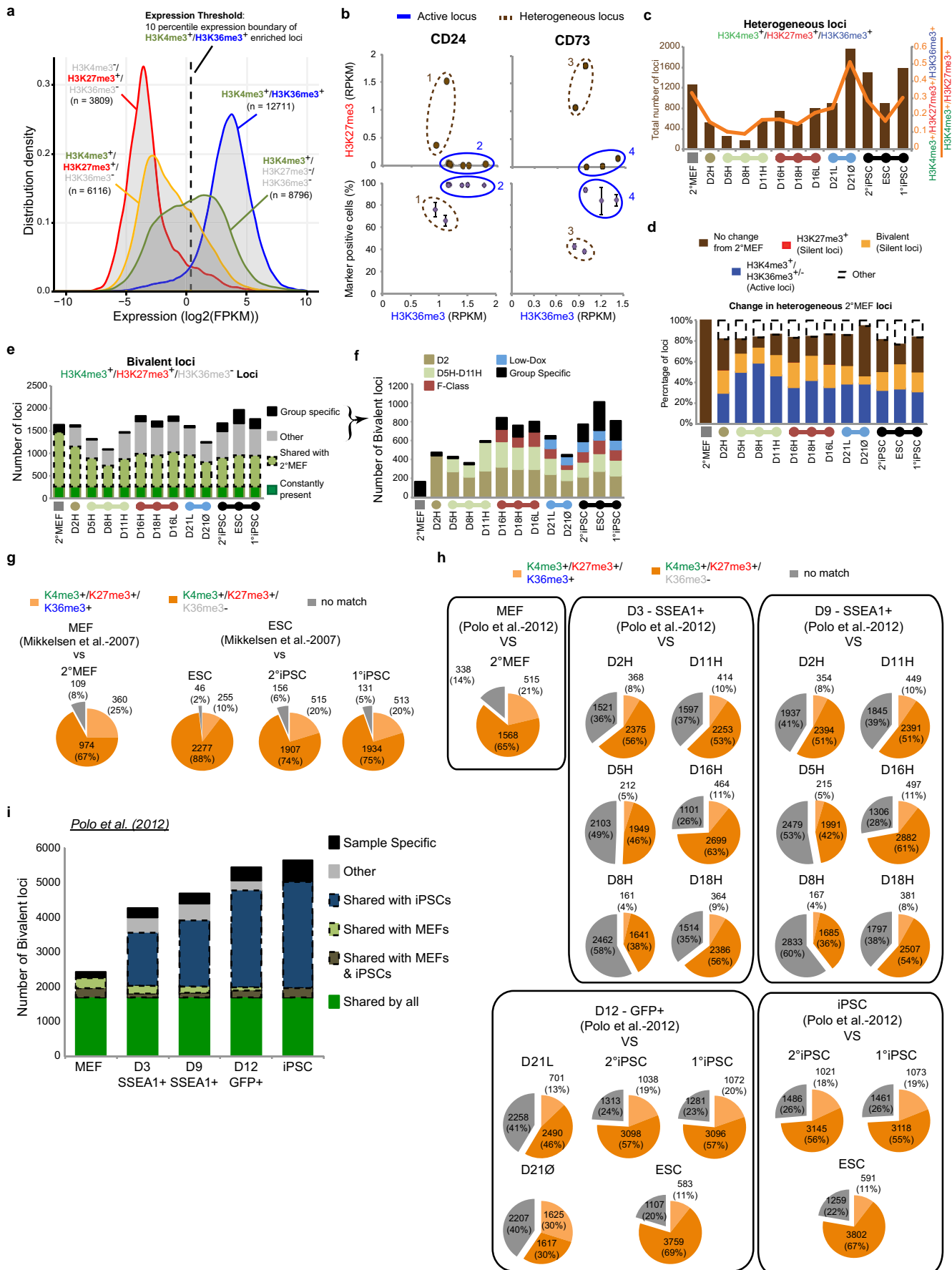
Extended Data Figure 8 | Determining expression threshold for defining bivalent loci and bivalency in other reprogramming systems. a, RNA-seq expression value ($\log_2$ of FPKM) distribution (as represented by density curves) of four categories of genes: (1) genes marked by H3K4me3 and H3K36me3 (blue line); (2) genes marked by H3K4me3 alone (green line); (3) genes marked by H3K27me3 alone (red line); and (4) genes marked by H3K4me3 and H3K27me3, but not H3K36me3 (orange line). Genes were combined from all the samples to identify each category. Expression threshold was defined as the 10th percentile expression boundary of genes marked by H3K4me3 and H3K36me3. Genes that were expressed at lower levels than this threshold were considered not expressed in subsequent analyses. **b**, Assessment of cellular heterogeneity in 1B reprogramming by chromatin mark and expression association of two cell surface markers: CD24 and CD73. Upper scatter plots show H3K27me3 versus H3K36me3 enrichment in individual samples. Lower plot shows percentage of cells expressing each marker for same samples as determined by FACs analysis. Active locus: $\text{H3K4me3}^{+}\text{H3K36me3}^{+}$

H3K27me3^{-}. Heterogeneous locus: $\text{H3K4me3}^{+}\text{H3K36me3}^{+}\text{H3K27me3}^{+}$. **c**, Absolute number (primary y axis) and proportion (secondary y axis) of false (heterogeneous) bivalent loci during secondary reprogramming. the presence of H3K36me3 distinguishes false bivalent loci ($\text{H3K4me3}^{+}\text{H3K27me3}^{+}$ H3K36me3^{+}) that represent heterogeneity from true bivalent loci that are transcriptionally repressed (H3K36me3^{-}). **d**, Tracking of histone mark status of secondary MEF heterogeneous loci. Heterogeneous loci resolve into silent and active loci during reprogramming. **e**, Total number of detected bivalent loci as defined by lack of H3K36me3 mark and expression levels below the threshold as shown in panel **a**. Dark and light green bar graphs highlight proportion shared among all samples and with secondary MEFs, respectively. **f**, Sequential addition of novel bivalent marks with respect to stages of reprogramming, as indicated by colours. **g, h**, Corresponding bivalent loci identified in 1B samples and two independent data sets[6,31]. **i**, Tracking of bivalent loci for Polo *et al.* reprogramming system[6]. For gene lists related to **e**, refer to Supplementary Table 2.
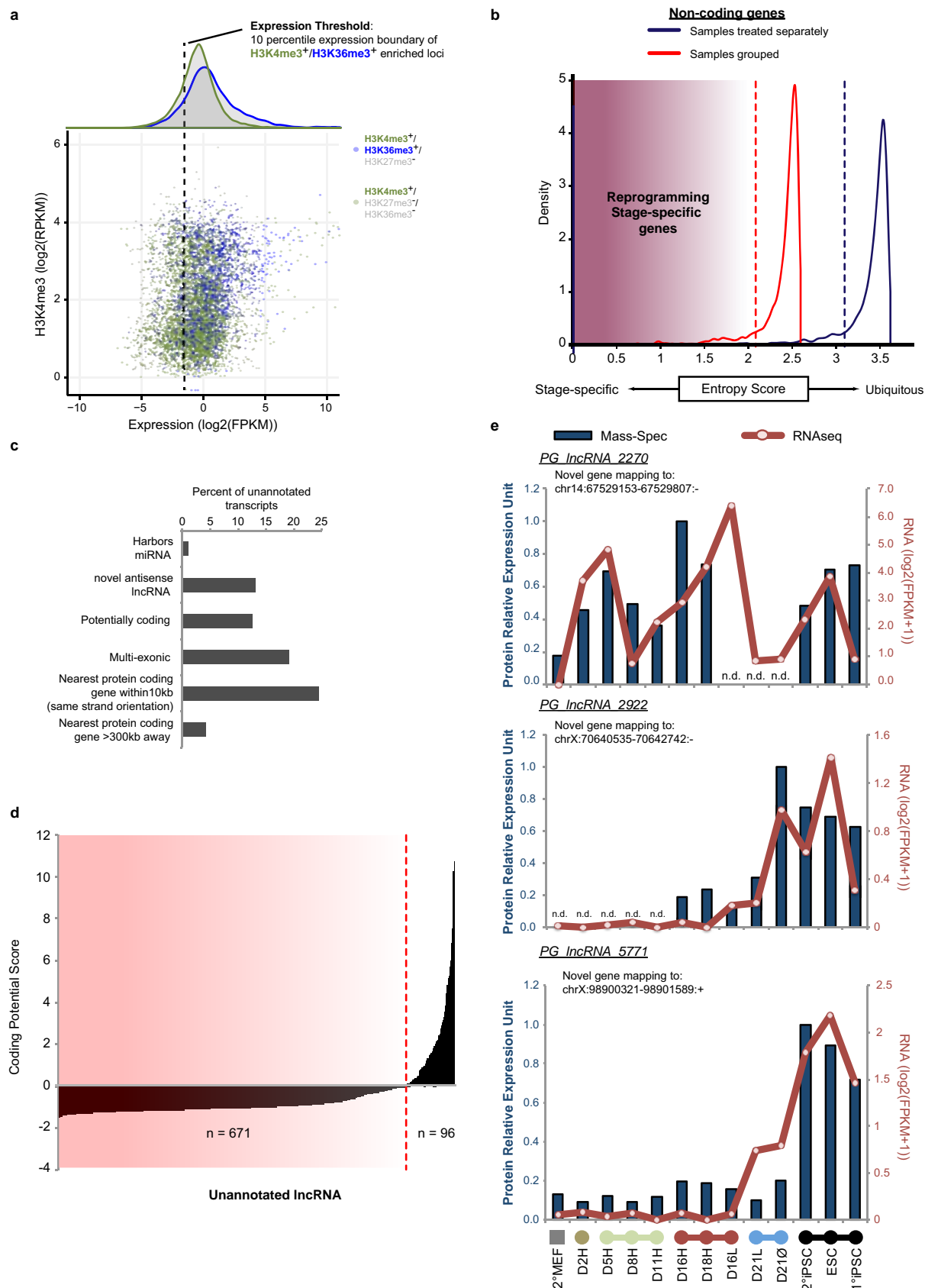
Extended Data Figure 9 | Long non-coding RNA expression analysis
a, Determination of expression threshold for lncRNA genes using H3K4me3 and H3K36me3 chromatin mark. **b**, Distribution of the entropy of non-coding gene expression for individual samples (blue) and sample groups (red) indicated as probability density curve. **c**, Percentage of unannotated transcripts with listed genomic features. **d**, Analysis of unannotated lncRNA transcripts for coding potential using coding potential calculator (CPC). (See Supplementary Information for details.) **e**, RNA and protein expression profiles of three novel coding transcripts.
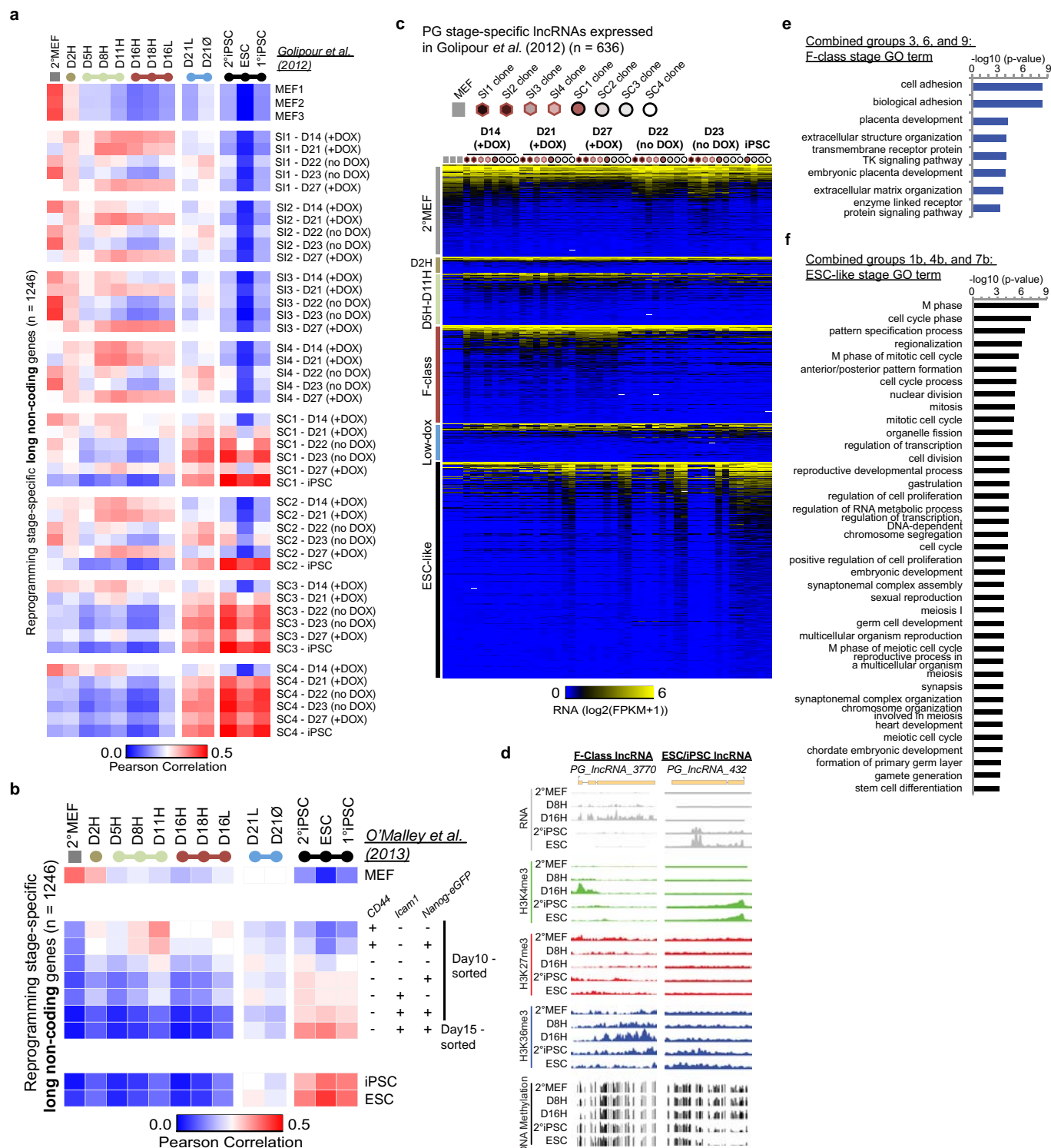
Extended Data Figure 10 | Comparison of lncRNA expression in 1B secondary reprogramming and other reprogramming systems. a, Pearson correlation analysis of differentially expressed un-annotated RNA transcripts for 1B reprogramming samples and secondary reprogramming clones that are competent or incompetent to become factor-independent secondary iPSCs (SC and SI clones, respectively)[7]. **b,** Pearson correlation analysis of differentially expressed unannotated RNA transcripts for 1B reprogramming samples and sorted reprogramming intermediates from ref. 8. **c,** Heat map of differentially expressed novel RNAs from 1B reprogramming samples with secondary reprogramming clones that are competent or incompetent to become

factor-independent secondary iPSCs (SC and SI clones, respectively)[7]. For gene lists related to **c**, refer to Supplementary Table 4. **d,** Read coverage histograms representing gene expression and epigenetic status of unannotated lncRNAs observed in F-class (D16H) and ESC-like state (secondary iPSCs). **e,** GO analysis results for genes downregulated in F-class state (FDR <1%), but unchanged in ESC-like state, from D8H (combined groups 3, 6 and 9). **f,** GO analysis results for genes upregulated in ESC-like state (FDR <1%), but unchanged in F-class state, from D8H (combined groups 1b, 4b and 7b). For gene lists, full GO term analyses and *P* values associated with **e, f** refer to Supplementary Table 5.

X-ray structure of a calcium-activated TMEM16 lipid scramblase

Janine D. Brunner[1], Novandy K. Lim[1], Stephan Schenck[1], Alessia Duerst[1] & Raimund Dutzler[1]

The TMEM16 family of proteins, also known as anoctamins, features a remarkable functional diversity. This family contains the long sought–after Ca²⁺-activated chloride channels as well as lipid scramblases and cation channels. Here we present the crystal structure of a TMEM16 family member from the fungus *Nectria haematococca* that operates as a Ca²⁺-activated lipid scramblase. Each subunit of the homodimeric protein contains ten transmembrane helices and a hydrophilic membrane-traversing cavity that is exposed to the lipid bilayer as a potential site of catalysis. This cavity harbours a conserved Ca²⁺-binding site located within the hydrophobic core of the membrane. Mutations of residues involved in Ca²⁺ coordination affect both lipid scrambling in *N. haematococca* TMEM16 and ion conduction in the Cl⁻ channel TMEM16A. The structure reveals the general architecture of the family and its mode of Ca²⁺ activation. It also provides insight into potential scrambling mechanisms and serves as a framework to unravel the conduction of ions in certain TMEM16 proteins.

The TMEM16 or anoctamin family constitutes a class of membrane proteins that is only expressed in eukaryotic organisms. In vertebrates the family encompasses ten members with high sequence conservation[1]. Despite their close relationship these proteins combine different functions as some members are Ca²⁺-activated ion channels while others work as Ca²⁺-activated scramblases[2], which catalyse the shuffling of lipids between the inner and outer leaflets of the bilayer in an ATP-independent manner. In 2008 three groups independently identified TMEM16A (or Ano1) as the long sought-after Ca²⁺-activated chloride channel (CaCC)[3–5]. After this discovery the name anoctamin was coined, synonymous for anion selectivity and the eight transmembrane spanning helices that were predicted by hydropathy analysis[3]. It has been shown that TMEM16A and TMEM16B (Ano2) share similar characteristics, although with different tissue distribution[3,5,6]. Whereas TMEM16A contributes to diverse physiological processes, such as epithelial chloride secretion, electrical signalling in smooth muscles and potentially also nociception[7,8], TMEM16B is expressed in the retina and in olfactory epithelia and might have a role in olfaction[6,9]. In further studies TMEM16F (Ano6) was shown to act as Ca²⁺-activated small-conductance cation channel[10], possibly also as Cl⁻ channel[11] and to have a role in Ca²⁺-activated lipid scrambling by facilitating the exchange of phosphatidylserine from the inner to the outer leaflet of the bilayer in blood platelets[10,12]. Similarly, TMEM16C, D, G and J (Ano3, 4, 7 and 9, respectively) were suggested to work as scramblases, although with variable lipid preference[13]. Recently a fungal TMEM16 homologue from *Aspergillus fumigatus* (afTMEM16) was found to scramble lipids as well after its purification and reconstitution into liposomes[14]. Besides its function as scramblase, afTMEM16 was also proposed to form non-selective ion channels with high conductance[14]. It is still a matter of debate how these closely related proteins can accommodate such a diversity of functional phenotypes[2,15].

Despite the functional breadth, characterized family members appear to share a similar mode of Ca²⁺ activation. This behaviour has been most thoroughly investigated for the chloride channel TMEM16A[3,16,17]. In TMEM16A Ca²⁺ activates the channel from the intracellular side at sub-micromolar concentrations with a half-maximum effective concentration (EC₅₀) that is voltage-dependent and decreases upon depolarization. Two conserved glutamate residues have been discovered to

be involved in the Ca²⁺ activation of ion conduction in TMEM16A[18,19] and TMEM16F[10] and scrambling in afTMEM16[14], thus indicating the conservation of this regulatory mechanism within the family.

Although we have by now gained considerable insight into the functional properties of certain family members, their architecture and its relation to mechanisms of action are still unknown. Here we present the X-ray structure of a TMEM16 homologue from *Nectria haematococca* (nhTMEM16). The dimeric protein shows a novel fold with ten membrane-spanning segments per subunit. The transmembrane domain contains a highly conserved region that is embedded within the hydrophobic core of the membrane comprising a Ca²⁺-binding site. Ca²⁺ binding by six residues, five of which carry a negative charge, controls the activation of scrambling in nhTMEM16 and ion conduction in TMEM16A. Our results thus have revealed a conserved structural framework that supports diverse functional properties within the family.

Functional characterization of nhTMEM16

To gain insight into the architecture of the TMEM16 family we screened 80 members in *Saccharomyces cerevisiae* and HEK tsA201 cells for over-expression and detergent stability, and were able to identify a homologue from the fungus *Nectria haematococca* (nhTMEM16), exhibiting the desired biochemical properties. The protein shares 48% of identical residues with the previously characterized afTMEM16 (with >70% homology within the transmembrane domain) (Extended Data Fig. 1a). Among mammalian proteins the relationship is closest to TMEM16H and K (Ano8 and 10) but it is still close to the more distantly related chloride channel TMEM16A (with homologies in the transmembrane region ranging between 39% and 42%), thus suggesting that all family members share a similar structural organization (Extended Data Fig. 1b,c). Unlike its mammalian counterparts, nhTMEM16 is not glycosylated. The solubilized protein is a dimer, as quantified by multi-angle light scattering, suggesting that the oligomeric organization is preserved in detergent solution (Extended Data Fig. 2a). To characterize its functional properties we have reconstituted the protein into liposomes and found, with respect to its scrambling activity, a very similar behaviour as described for the related afTMEM16. The function as lipid scramblase

[1]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.
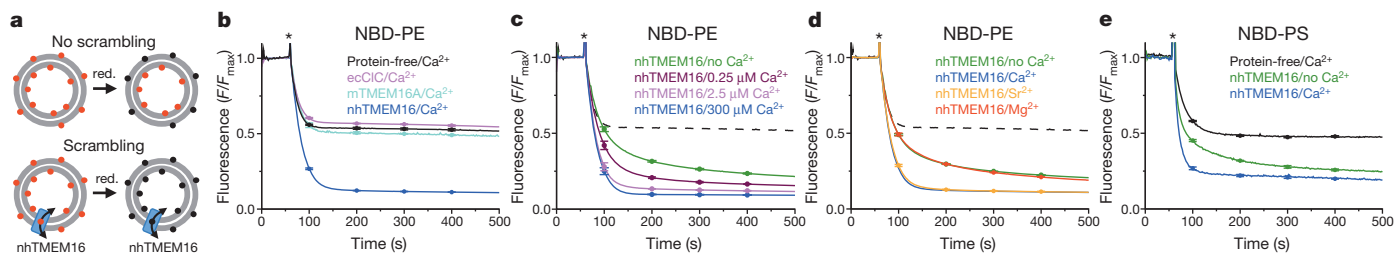
Figure 1 | Phospholipid scrambling by nhTMEM16. a, Scheme of the assay depicting the reduction of NBD-labelled phospholipids in the outer leaflet of the bilayer upon addition of dithionite (red.). **b**, Scrambling of NBD-PE. Traces of protein-free liposomes and proteoliposomes containing either nhTMEM16, the *E. coli* ClC transporter (ecClC) or mTMEM16A are shown. Asterisk marks addition of dithionite. **c**, Ca^{2+} dependence of NBD-PE scrambling by nhTMEM16. **d**, Influence of other divalent cations on NBD-PE scrambling by nhTMEM16. In **c** and **d**, protein-free liposomes (dashed lines) are shown for comparison. **e**, Scrambling of liposomes containing NBD-PS. Traces of protein-free liposomes and proteoliposomes containing nhTMEM16 in the absence and presence of Ca^{2+} are shown. **b–e**, Traces show averages of 3–4 measurements, standard deviations are included for selected time points. Unless stated otherwise, solutions contain 0.3 mM of the indicated free divalent cations. Scrambling experiments were replicated three times with similar results.

was investigated by an assay that monitors the reduction of fluorescently labelled lipids by sodium dithionite on the outer leaflet of liposomes[14,20] (Fig. 1a, Extended Data Fig. 2b–d). Our results demonstrate that nhTMEM16 catalyses the movement of nitrobenzoxadiazole-phosphatidylethanolamine (NBD-PE) and NBD-phosphatidylserine (NBD-PS) between the two layers of the liposome membrane (Fig. 1b, e). The observed effect is not due to permeation of dithionite through the protein (Extended Data Fig. 2b). Furthermore, we found that this catalytic function is enhanced by Ca^{2+} at submicromolar concentrations (Fig. 1c). Besides Ca^{2+}, we also observed activation for Sr^{2+} but not for Mg^{2+} (Fig. 1d). Scrambling in nhTMEM16 containing proteoliposomes measured under Ca^{2+}-free conditions may either be due to constitutive activity of the ligand-free scramblase or originate from traces of Ca^{2+} still bound to the protein (Supplementary Discussion, Extended Data Fig. 2d). To investigate whether nhTMEM16 would also function as ion channel, we have attempted to study ion conduction from proteoliposomes fused to artificial lipid bilayers and by patch-clamp electrophysiology of HEK293T cells expressing the protein. However, in neither case did we find any convincing evidence for ion channel activity (Extended Data Fig. 3 and Supplementary Discussion).

nhTMEM16 structure

For structure determination, nhTMEM16 was crystallized in two different crystal forms (CF1 and CF2), each containing a dimer in the asymmetric unit, for which we have collected data at 3.3 and 3.4 Å resolution respectively (Extended Data Fig. 4a). Initial phases, obtained by the Se-Met single-wavelength anomalous dispersion method, were improved and extended by non-crystallographic symmetry and cross-crystal averaging. The resulting electron density was of high quality and allowed the unambiguous interpretation by an atomic model (Extended Data Figs 4 and 5). The structure of the dimer is depicted in Fig. 2a. Both subunits are related by twofold symmetry and show very similar conformations. When viewed from the extracellular side the dimer has a rhombus-like shape with about 130 Å in the long and 40 Å in the short dimension (Extended Data Fig. 6). The topology of the nhTMEM16 subunit is shown in Fig. 2b. Both termini are structured and located on the cytoplasmic side of the membrane. The α-helices and β-strands of the amino-terminal domain are organized in a ferredoxin-like fold. The three α-helices of the carboxy terminus are wrapped around the N-terminal domain of the adjacent subunit, thereby constituting a large part of the subunit interface (Fig. 2c). The transmembrane domain starts with two short α-helices (α0a and α0b), followed by ten membrane-spanning segments (α1–α10). The two initial helices form a hairpin with amphiphilic properties, with its hydrophobic side interacting with α-helices 1 and 8. A model of the protein embedded in a lipid bilayer suggests that both helices only peripherally interact with the inner leaflet of the membrane (Extended Data Fig. 7a, b). Helices α1–α10, in contrast, all traverse the entire membrane with some of them being bent and tilted
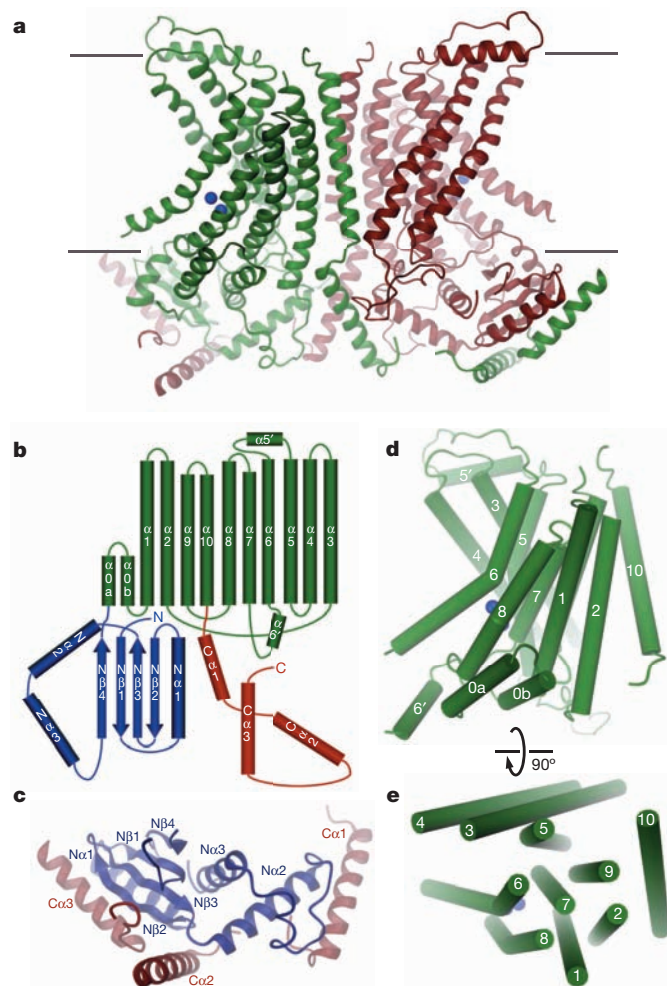
Figure 2 | nhTMEM16 structure. a, Ribbon representation of the nhTMEM16 dimer. The view is from within the membrane. Bound calcium ions are shown as blue spheres. The membrane boundary is indicated. **b**, Topology of the nhTMEM16 subunit. The transmembrane domain is coloured in green, the N- and C-terminal domains in blue and red, respectively. **c**, View of the cytoplasmic domains. The interaction of the N-terminal domain with the C-terminal domain of the adjacent subunit is shown. **d**, Transmembrane domain with α-helices shown as cylinders and labelled. The view is as in **a**. **e**, Organization of transmembrane helices. The view is from the extracellular side. Loop regions are omitted for clarity. Figures 2–4 and 6 were prepared with DINO (http://www.dino3d.org/) and show the structure determined in crystal form 1 (CF1) unless stated otherwise.

with respect to its plane (Fig. 2d, e). The transmembrane segments are connected by loop regions of variable length, two of which contain short helical regions (named α5′ and α6′, according to the preceding transmembrane region), on the extracellular and cytoplasmic side, respectively. The arrangement of α-helices does not follow any obvious symmetry or show any relationship to known membrane protein structures.

Dimer interface and dimer cavity

The dimeric organization of nhTMEM16 is reflected in the extended interface between the two subunits, which buries 9,650 Å^2 of the combined molecular surface. The largest part of this interface is contributed by interactions between the N- and C-terminal domains, whereas the contact area of 1,520 Å^2 between the transmembrane domains is comparably small. In the transmembrane region the dimer interface is formed by interactions of residues in the N-terminal part of α-helix 10 at the extracellular side close to the symmetry axis and interactions between α-helices 3 and 10 at their cytoplasmic end (Fig. 3a, b). Mutual interactions between residues of α-helices 10 involve hydrophobic contacts and a pair of salt bridges between a glutamate and a histidine side-chain that are conserved within the family, except for TMEM16A and B (Extended Data Fig. 7c). The arrangement of helices close to the dimer interface generates a large pore-like structure across the transmembrane region, the dimer cavity, which contains two separate 15 Å wide entrances at the extracellular side and which merges to one big, about 30 Å wide vestibule, at the intracellular half of the membrane (Fig. 3b, Extended Data Fig. 7d, e). Although on the inside this large cavity is confined by residues of the cytoplasmic domains, several fenestrations create access to the cytoplasm. In the transmembrane region the dimer cavity is accessible to the outer leaflet of the membrane via two v-shaped gaps framed by α-helices 3 and 10 from adjacent subunits (Fig. 3b, Extended Data Fig. 7e). Within the membrane, the vestibule is predominantly composed of hydrophobic and aromatic residues, which are conserved within the protein family, whereas there are several polar and charged residues found at the intracellular part outside of the predicted membrane region (Extended Data Fig. 7d, e). We thus suppose that the dimer cavity may be packed with lipids. The inside of this large cavity contains excess electron density, which is, however, not sufficiently ordered to be attributed to either solvent, or detergent and lipids (Extended Data Fig. 7f). It is currently not clear whether this region has a critical role for protein function.

Subunit cavity and Ca^{2+}-binding region

With respect to function, the most remarkable feature is found on the surface opposite to the dimer interface. Here a narrow crevice that spans the entire membrane is formed by α-helices 3–7 of the same subunit

(Fig. 3c, Extended Data Fig. 8). These α-helices surround an 8–11 Å wide cavity that is twisted like a 'spiral staircase' and on one side is exposed to the membrane. In contrast to the dimer cavity the surface of the subunit cavity is strongly hydrophilic, despite its exposure to the lipid environment (Fig. 3c). Furthermore, it harbours residues for which equivalent positions have previously been shown to be involved in ion conduction in TMEM16A[18] and to influence ion selectivity in TMEM16A and F[10]. These observations make this region a prime candidate for the translocation path in both channels and scramblases. Within the hydrophobic core of the membrane, at a distance corresponding to about one third of its thickness from the intracellular side, the subunit cavity is lined by residues of α-helices 6 and 7, which are part of a conserved Ca^{2+}-binding site (Fig. 4a, b and Extended Data Fig. 8d). In the crystal structure, we have detected bound Ca^{2+} ions by anomalous scattering. Two peaks of strong anomalous difference density were found in each subunit at equivalent places (Fig. 4c, d). These peaks are separated by a distance of 4.2 Å and they are surrounded by three glutamates, two aspartates and an asparagine located on α-helices 6, 7 and 8 (Fig. 4e). Although from our data we cannot tell with certainty whether one or two Ca^{2+} ions are bound at the same time, simultaneous occupancy seems possible owing to the high density of negative charge in this region. All residues involved in Ca^{2+} binding are highly conserved within the TMEM16 family, which strongly supports a common Ca^{2+} binding and activation mode.

Functional investigation of the Ca^{2+}-binding region

Since the nhTMEM16 structure has allowed the identification of a conserved Ca^{2+}-binding site, we were interested to investigate the relevance of these interactions for activation in nhTMEM16 and TMEM16A. The two glutamates located in α-helix 7 have previously been shown to play an important role in the activation of TMEM16A[18], F[10] and afTMEM16[14] by Ca^{2+}. A Ca^{2+}-binding site triple-mutant of nhTMEM16 combining mutations of residues involved in Ca^{2+} coordination identified in this study (that is, E452Q, E535Q and D539N) shows only weak scrambling activity that is not enhanced by Ca^{2+} (Fig. 5a). To probe the importance of the same residues for the activation of ion conduction in murine TMEM16A (mTMEM16A), we have expressed the protein in HEK293T cells and monitored the Ca^{2+}-dependence in binding site mutants by patch-clamp electrophysiology. As previously shown Ca^{2+} activates mTMEM16A in a voltage-dependent manner with an apparent affinity that is higher at positive than at negative potentials[3] (Fig. 5b). At 80 mV Ca^{2+} activates the wild-type protein with an EC$_{50}$ of 0.36 μM and a Hill coefficient of 2.3, which indicates a cooperative process that involves the binding of more than one ion. Single mutants of each residue contributing to the observed interactions in the Ca^{2+}-binding region
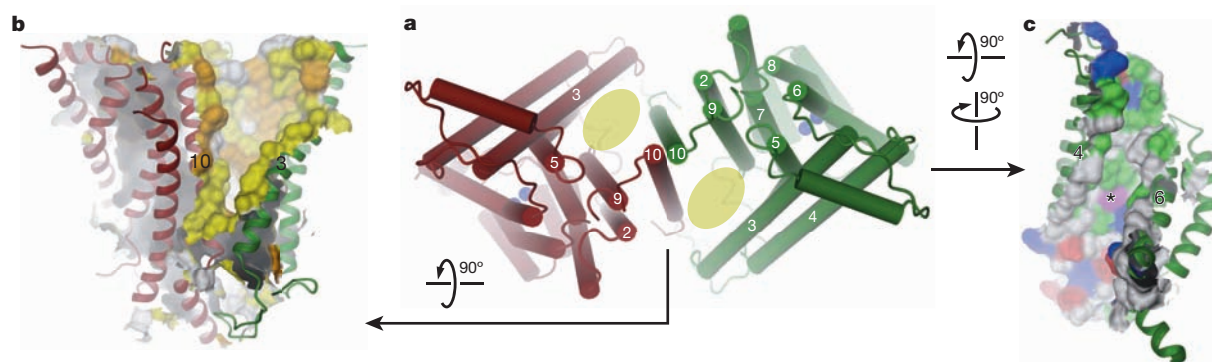
Figure 3 | Dimer interface and subunit cavity. a, Transmembrane domain of the nhTMEM16 dimer viewed from the extracellular side with yellow ovals indicating the location of the dimer cavity. Helices are represented as cylinders. **b,** Dimer cavity viewed from within the membrane. Helices of both subunits lining the cavity are displayed. The solvent-accessible surface of the cavity is shown with locations of hydrophobic and aromatic residues coloured in yellow and orange, respectively. **c,** Subunit cavity viewed from within the membrane. The solvent-accessible surface is coloured according to the properties of contacting residues (red, acidic; blue, basic; green, polar). A position that was shown to influence the ion selectivity in TMEM16A and TMEM16F is coloured in magenta and labelled with an asterisk.
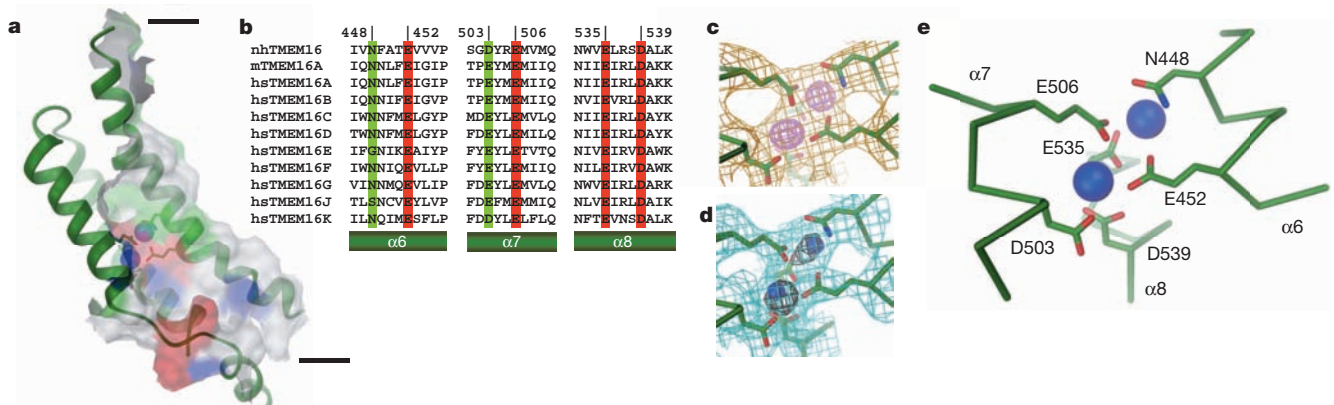
Figure 4 | Ca²⁺-binding site. a, Location of the Ca²⁺-binding site in relation to the subunit cavity. The view is from within the membrane, the colour coding as in Fig. 3c. **b**, Sequence alignment. Conserved amino acids of the Ca²⁺-binding site are highlighted in red (identical) and green (homologous). hs, *Homo sapiens*. **c**, View of the Ca²⁺-binding site in the refined structure of CF2. The $2F_o - F_c$ electron density (at 3.5 Å, contoured at 1σ, orange) is shown

superimposed on the model. Anomalous difference electron density (at 3.8 Å, contoured at 5σ) is shown in magenta. **d**, View of the Ca²⁺-binding site in the refined structure of CF1. The $2F_o - F_c$ electron density (at 3.3 Å, 1σ, cyan, 5σ, black) is shown superimposed on the model containing Ca²⁺ ions (blue spheres). **e**, Model of the Ca²⁺-binding site.

shift the EC₅₀ to higher Ca²⁺ concentrations (Fig. 5c and d, Extended Data Figs 9 and 10). The strongest effect was observed for Glu 654, where we have not observed any activation for E654Q and only low currents at high Ca²⁺ concentration for E654A, despite the strong plasma membrane expression of the channel (Extended Data Fig. 10h). Similar results were reported in a recent study that was based on the mutation of conserved acidic residues[21]. Taken together, our functional experiments on nhTMEM16 and mTMEM16A suggest that Ca²⁺ binding by equivalent residues regulates both functional branches of the family by a common mechanism.

Discussion

The structure of nhTMEM16 has revealed a framework for the TMEM16/anoctamin family. Whereas its homodimeric organization is consistent with previous investigations of TMEM16A, B, F and afTMEM16[14,22–24], a direct interaction between the N termini, which was proposed to be involved in dimerization of TMEM16A, is not observed[24]. In nhTMEM16, each subunit contains ten membrane-spanning helices, which differs from the eight transmembrane segments predicted from hydropathy analysis[3,25] (Extended Data Fig. 1c). It is however noteworthy that a recent study, which has revised the originally proposed topology, has correctly identified residues of the Ca²⁺-binding site and the extracellular entry to the putative pore region[18]. The structure harbours two regions that are presumably in contact with the membrane; the dimer cavity, a large and predominantly hydrophobic structure at the dimer interface, and the subunit cavity, a hydrophilic membrane-spanning

crevice contained within each subunit that resembles a spiral staircase. Whereas the functional relevance of the dimer cavity is currently unclear, the subunit cavity is linked to Ca²⁺ activation and probably also to catalytic properties of the protein (Fig. 3, Extended Data Fig. 8).

As a scramblase nhTMEM16 has provided first structural insight into an important class of transport molecules. These proteins catalyse the passive movement of lipids between the two leaflets of a bilayer, a process that is essential for membrane biogenesis in the endoplasmic reticulum[26] and the shuffling of lipids in several processes, including blood coagulation[10,12], apoptosis[27], glycosylation[28] and the assembly of the bacterial cell wall[29]. To lower the large intrinsic energy barrier associated with lipid flipping, it was proposed that scramblases would provide a hydrophilic path to facilitate the movement of the polar headgroups across the bilayer[26,30]. The subunit cavity of nhTMEM16 would ideally meet these requirements, as it is hydrophilic, accessible to the membrane and of sufficient dimensions to accommodate a phospholipid headgroup (Fig. 6a). These properties would also be consistent with the broad range of lipids that have been shown to be translocated by afTMEM16[14] and other Ca²⁺-activated lipid scramblases[31]. Proteins with scrambling function belong to various families[13,20,27,29], some of which are still disputed[32] or have not yet been identified on a molecular level. It will thus be interesting to see to which extent structural and mechanistic features proposed in our work are shared by unrelated scramblases and whether still unassigned scrambling processes would be catalysed by TMEM16 proteins.

In TMEM16A, B and F, the subunit cavity may constitute the ion conduction pore, as suggested by the influence of point mutations of
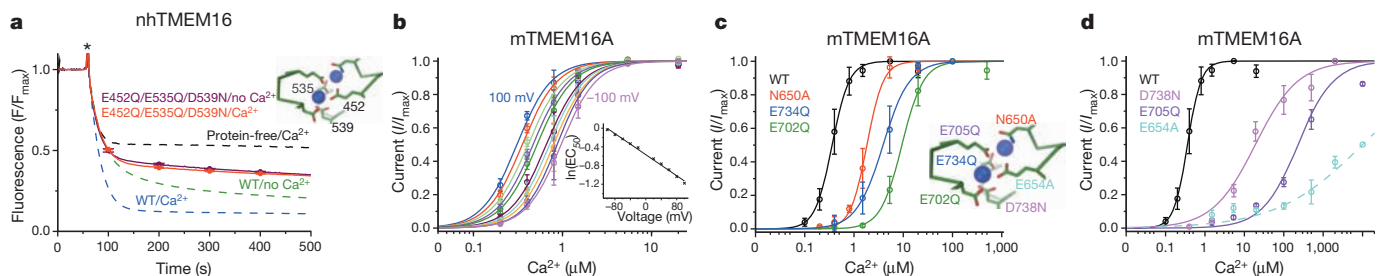
Figure 5 | Functional properties of Ca²⁺-binding site mutants. a, Scrambling of the nhTMEM16 mutant E452Q/E535Q/D539N. Data of proteoliposomes containing NBD-PE and the triple mutant in absence and presence of 0.3 mM free Ca²⁺ are shown. Traces and statistics are as in Fig. 1 with dashed lines shown for comparison. Inset indicates the position of the mutations. WT, wild type. **b**, Voltage dependence of Ca²⁺ activation in mTMEM16A measured from excised inside-out patches. Currents were

normalized to the maximum, lines show a fit to a Hill equation. The voltage dependence is shown as inset. **c, d**, Activation of Ca²⁺-binding site mutants. Data were recorded at 80 mV and normalized to the maximum. Lines show a fit to a Hill equation (the fit for E654A was estimated). Numbers in inset correspond to mTMEM16A. Data in **b–d** are averages from 3–4 independent measurements, errors are s.d.
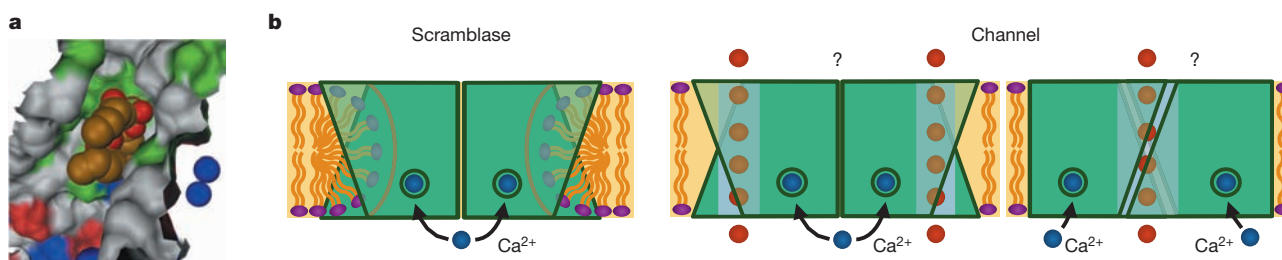
Figure 6 | Mechanism. a, Model of phosphatidylcholine in the subunit cavity. The acyl chains have been truncated for clarity. Bound Ca^{2+} ions are shown as blue spheres. **b**, Scheme illustrating the potential relationship between scramblases and ion channels within the TMEM16 family. Two distinct dimeric arrangements of TMEM16 channels are shown with one resembling the nhTMEM16 structure (centre) and the other formed by monomers that interact via the subunit cavities (right).

residues facing the subunit cavity on ion selectivity and conductance[10,18]. In TMEM16A and B, it is still puzzling how the subunit cavity might provide the aqueous environment required for ion conduction and, conversely, how lipid scrambling would be precluded in these proteins[13,14,33] (Fig. 1b). This distinction could be accomplished by local structural differences in an assembly as observed in nhTMEM16. Alternatively, it is tempting to speculate that in TMEM16 channels, the monomers might be turned by 180° and interact via their subunit cavities to form an enclosed pore. These different dimer assemblies would provide a plausible explanation for the functional dichotomy in TMEM16 proteins and are largely compatible with the structure of the monomer (Fig. 6b). Whereas the arrangement that is similar to nhTMEM16 would contain two potentially independent pores, as seen in the ClC family[34], the latter would probably result in a single ion conduction path.

For scramblases it will be important to investigate whether lipid movement locally distorts the membrane and whether ions can pass through the bilayer by binding to the polar headgroups of lipids that are in the process of being scrambled, which could potentially give rise to the small ion conductance observed in TMEM16F[10]. The large channels observed in afTMEM16[14] may be distinct from the described process and no such property has so far been detected for the closely related homologue investigated in this study. Similarly, a potential chloride-selective ion conductance of TMEM16F[11], which was proposed to be independent of its role in scrambling[35] and a similar function in other TMEM16 proteins[36] still requires closer investigation. The coexistence of ion channels and lipid scramblases in the TMEM16 family is somewhat reminiscent of P-type ATPases[37], which include primary active ion pumps and ATP-driven lipid flippases with similar molecular architecture. Also in this family both transport functions were proposed to localize in the same area. In flippases polar headgroups are thought to interact with a hydrophilic groove during bilayer passage[38], thereby facing a similar environment to that found in the subunit cavity of nhTMEM16.

Despite the breadth of functional behaviours, both branches of the TMEM16 family share the mechanism by which their activity is regulated by Ca^{2+}. The nhTMEM16 structure reveals a conserved Ca^{2+}-binding site contained within each subunit that is positioned within the hydrophobic core of the bilayer in proximity to the subunit cavity. Although this site potentially harbours two Ca^{2+} ions, it is currently not known whether binding of one or two ions is required to activate the protein. The location of this region within the membrane provides an explanation for the voltage-dependence of Ca^{2+} activation observed in TMEM16A[3,5], B[39] and TMEM16F[10]. This effect probably originates from the fact that the ion has to cross a fraction of the transmembrane electric field to reach the binding site[40], a model that was already proposed in an early study on CaCC activation[41]. While in our structure Ca^{2+} ions are buried within the protein, their entry from the cytoplasm via the subunit cavity or another path that becomes accessible in the ligand-free protein appears plausible (Fig. 4a, Extended Data Fig. 8). Ca^{2+} binding may either induce a conformational change in the protein that underlies its activation or modify the electrostatics in the close-by

subunit cavity and in that way regulate the conductive properties of this region. Although the mechanism described is probably common for this protein family, there may be additional modes of regulation in certain TMEM16 proteins[42–46]. Our study has shed light on the unique properties of the TMEM16 protein family that does not resemble known classes of membrane proteins with respect to structure nor function. While detailed mechanisms of action are still unknown, the structure of nhTMEM16 has provided a template that will guide the future investigation of structure–function relationships.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Milenkovic, V. M., Brockmann, M., Stohr, H., Weber, B. H. & Strauss, O. Evolution and functional divergence of the anoctamin family of membrane proteins. *BMC Evol. Biol.* **10,** 319 (2010).
2. Pedemonte, N. & Galietta, L. J. Structure and function of TMEM16 proteins (anoctamins). *Physiol. Rev.* **94,** 419–459 (2014).
3. Yang, Y. D. *et al.* TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature* **455,** 1210–1215 (2008).
4. Caputo, A. *et al.* TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science* **322,** 590–594 (2008).
5. Schroeder, B. C., Cheng, T., Jan, Y. N. & Jan, L. Y. Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. *Cell* **134,** 1019–1029 (2008).
6. Pifferi, S., Cenedese, V. & Menini, A. Anoctamin 2/TMEM16B: a calcium-activated chloride channel in olfactory transduction. *Exp. Physiol.* **97,** 193–199 (2012).
7. Ferrera, L., Zegarra-Moran, O. & Galietta, L. J. Ca^{2+}-activated Cl^- channels. *Compr. Physiol.* **1,** 2155–2174 (2011).
8. Huang, F., Wong, X. & Jan, L. Y. International Union of Basic and Clinical Pharmacology. LXXXV: calcium-activated chloride channels. *Pharmacol. Rev.* **64,** 1–15 (2012).
9. Billig, G. M., Pal, B., Fidzinski, P. & Jentsch, T. J. Ca^{2+}-activated Cl^- currents are dispensable for olfaction. *Nature Neurosci.* **14,** 763–769 (2011).
10. Yang, H. *et al.* TMEM16F forms a Ca^{2+}-activated cation channel required for lipid scrambling in platelets during blood coagulation. *Cell* **151,** 111–122 (2012).
11. Martins, J. R. *et al.* Anoctamin 6 is an essential component of the outwardly rectifying chloride channel. *Proc. Natl Acad. Sci. USA* **108,** 18168–18172 (2011).
12. Suzuki, J., Umeda, M., Sims, P. J. & Nagata, S. Calcium-dependent phospholipid scrambling by TMEM16F. *Nature* **468,** 834–838 (2010).
13. Suzuki, J. *et al.* Calcium-dependent phospholipid scramblase activity of TMEM16 protein family members. *J. Biol. Chem.* **288,** 13305–13316 (2013).
14. Malvezzi, M. *et al.* Ca^{2+}-dependent phospholipid scrambling by a reconstituted TMEM16 ion channel. *Nature Commun.* **4,** 2367 (2013).
15. Kunzelmann, K. *et al.* Molecular functions of anoctamin 6 (TMEM16F): a chloride channel, cation channel, or phospholipid scramblase? *Pflugers Arch.* **466,** 407–414 (2014).
16. Kuruma, A. & Hartzell, H. C. Bimodal control of a Ca^{2+}-activated Cl^- channel by different Ca^{2+} signals. *J. Gen. Physiol.* **115,** 59–80 (2000).
17. Ni, Y. L., Kuan, A. S. & Chen, T. Y. Activation and inhibition of TMEM16A calcium-activated chloride channels. *PLoS ONE* **9,** e86734 (2014).
18. Yu, K., Duran, C., Qu, Z., Cui, Y. Y. & Hartzell, H. C. Explaining calcium-dependent gating of anoctamin-1 chloride channels requires a revised topology. *Circ. Res.* **110,** 990–999 (2012).
19. Terashima, H., Picollo, A. & Accardi, A. Purified TMEM16A is sufficient to form Ca^{2+}-activated Cl^- channels. *Proc. Natl Acad. Sci. USA* **110,** 19354–19359 (2013).
20. Menon, I. *et al.* Opsin is a phospholipid flippase. *Curr. Biol.* **21,** 149–153 (2011).
21. Tien, J. *et al.* A comprehensive search for calcium binding sites critical for TMEM16A calcium-activated chloride channel activity. *eLife* **3,** e02772 (2014).
22. Fallah, G. *et al.* TMEM16A(a)/anoctamin-1 shares a homodimeric architecture with CLC chloride channels. *Mol. Cell. Proteomics* **10,** M110.004697 (2011).

23. Sheridan, J. T. *et al.* Characterization of the oligomeric structure of the Ca^{2+}-activated Cl$^-$ channel Ano1/TMEM16A. *J. Biol. Chem.* **286,** 1381–1388 (2011).
24. Tien, J., Lee, H. Y., Minor, D. L. Jr, Jan, Y. N. & Jan, L. Y. Identification of a dimerization domain in the TMEM16A calcium-activated chloride channel (CaCC). *Proc. Natl Acad. Sci. USA* **110,** 6352–6357 (2013).
25. Hartzell, H. C., Yu, K., Xiao, Q., Chien, L. T. & Qu, Z. Anoctamin/TMEM16 family members are Ca^{2+}-activated Cl$^-$ channels. *J. Physiol. (Lond.)* **587,** 2127–2139 (2009).
26. Sanyal, S. & Menon, A. K. Flipping lipids: why an' what's the reason for? *ACS Chem. Biol.* **4,** 895–909 (2009).
27. Suzuki, J., Denning, D. P., Imanishi, E., Horvitz, H. R. & Nagata, S. Xk-related protein 8 and CED-8 promote phosphatidylserine exposure in apoptotic cells. *Science* **341,** 403–406 (2013).
28. Sanyal, S. & Menon, A. K. Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proc. Natl Acad. Sci. USA* **107,** 11289–11294 (2010).
29. Mohammadi, T. *et al.* Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J.* **30,** 1425–1432 (2011).
30. Zwaal, R. F., Comfurius, P. & Bevers, E. M. Scott syndrome, a bleeding disorder caused by defective scrambling of membrane phospholipids. *Biochim. Biophys. Acta* **1636,** 119–128 (2004).
31. Dekkers, D. W., Comfurius, P., Bevers, E. M. & Zwaal, R. F. Comparison between Ca^{2+}-induced scrambling of various fluorescently labelled lipid analogues in red blood cells. *Biochem. J.* **362,** 741–747 (2002).
32. Sham, L. T. *et al.* Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* **345,** 220–222 (2014).
33. Suzuki, T., Suzuki, J. & Nagata, S. Functional swapping between transmembrane proteins TMEM16A and TMEM16F. *J. Biol. Chem.* **289,** 7438–7447 (2014).
34. Miller, C. Open-state substructure of single chloride channels from *Torpedo* electroplax. *Phil. Trans. R. Soc. Lond. B* **299,** 401–411 (1982).
35. Kmit, A. *et al.* Calcium-activated and apoptotic phospholipid scrambling induced by Ano6 can occur independently of Ano6 ion currents. *Cell Death Dis.* **4,** e611 (2013).
36. Tian, Y., Schreiber, R. & Kunzelmann, K. Anoctamins are a family of Ca^{2+}-activated Cl$^-$ channels. *J. Cell Sci.* **125,** 4991–4998 (2012).
37. Baldridge, R. D. & Graham, T. R. Identification of residues defining phospholipid flippase substrate specificity of type IV P-type ATPases. *Proc. Natl Acad. Sci. USA* **109,** E290–E298 (2012).
38. Vestergaard, A. L. *et al.* Critical roles of isoleucine-364 and adjacent residues in a hydrophobic gate control of phospholipid transport by the mammalian P4-ATPase ATP8A2. *Proc. Natl Acad. Sci. USA* **111,** E1334–E1343 (2014).
39. Pifferi, S., Dibattista, M. & Menini, A. TMEM16B induces chloride currents activated by calcium in mammalian cells. *Pflugers Arch.* **458,** 1023–1038 (2009).
40. Woodhull, A. M. Ionic blockage of sodium channels in nerve. *J. Gen. Physiol.* **61,** 687–708 (1973).
41. Arreola, J., Melvin, J. E. & Begenisich, T. Activation of calcium-dependent chloride channels in rat parotid acinar cells. *J. Gen. Physiol.* **108,** 35–47 (1996).
42. Xiao, Q. *et al.* Voltage- and calcium-dependent gating of TMEM16A/Ano1 chloride channels are physically coupled by the first intracellular loop. *Proc. Natl Acad. Sci. USA* **108,** 8891–8896 (2011).
43. Vocke, K. *et al.* Calmodulin-dependent activation and inactivation of anoctamin calcium-gated chloride channels. *J. Gen. Physiol.* **142,** 381–404 (2013).
44. Yu, K., Zhu, J., Qu, Z., Cui, Y. Y. & Hartzell, H. C. Activation of the Ano1 (TMEM16A) chloride channel by calcium is not mediated by calmodulin. *J. Gen. Physiol.* **143,** 253–267 (2014).
45. Tian, Y. *et al.* Calmodulin-dependent activation of the epithelial calcium-dependent chloride channel TMEM16A. *FASEB J.* **25,** 3 1058–1068 (2011).
46. Jung, J. *et al.* Dynamic modulation of ANO1/TMEM16A HCO$_3^-$ permeability by Ca^{2+}/calmodulin. *Proc. Natl Acad. Sci. USA* **110,** 360–365 (2013).

Author Contributions J.D.B. screened homologues, purified and crystallized nhTMEM16 and performed scrambling experiments. N.K.L. screened and crystallized homologues, performed electrophysiological recordings and did the MALS measurement. S.S. started the project, made expression vectors and aided in cell culture. A.D. screened homologues. R.D. assisted J.D.B. and N.K.L. during structure determination. J.D.B., N.K.L., S.S. and R.D. jointly planned experiments, analysed data and wrote the manuscript.

METHODS

Cloning. The gene encoding nhTMEM16 from *Nectria haematococca* (PubMed accession number XM_003045982) was synthesized by GenScript and the gene encoding murine TMEM16A (mTMEM16A, isoform a) was obtained from Ima-Genes (Clone IRAVp968B10135D). Expression vectors were modified to be compatible with FX-cloning[47]. For expression in *S. cerevisiae* nhTMEM16 was cloned into a modified pYES2/CT plasmid (Life Technologies) as C-terminal fusion to a cassette encoding EGFP, preceded by a His[10]-tag and followed by a HRV 3C cleavage site (crystallization construct) or as N-terminal fusion to a cassette containing a streptavidin-binding peptide (SBP) tag[48] preceded by a Myc tag and a HRV 3C cleavage site (scramblase assay construct). For expression in tsA201 cells, nhTMEM16 and mTMEM16A were cloned into a modified pcDNA3.1 vector (Invitrogen), bearing a 5′ UTR (untranslated region) of hVEGF (from pcDNA4/HisMax, Invitrogen) upstream of the start codon. nhTMEM16 as well as mTMEM16A (isoform a) contained a C-terminal HRV 3C cleavage site, a Myc- and an SBP tag (scramblase assay constructs). For expression in HEK293T cells mTMEM16A (isoform ac) as well as nhTMEM16 were expressed with a C-terminal fusion encoding a Venus-yellow fluorescent protein (YFP)[49], a Myc- and an SBP tag[50] (defined as mTMEM16A–YFP, nhTMEM16–YFP; used in electrophysiological recordings). The mTMEM16Aac isoform used in patch-clamp experiments was generated by PCR. All point mutations were introduced by site-directed mutagenesis.

Protein expression. For expression of nhTMEM16 and its mutants, the pYES2/CT vectors carrying the respective genes were transformed into *S. cerevisiae* FGY217 cells carrying an *URA* deletion for positive selection as described[51]. Cells were grown at 30 °C in fermentation culture in yeast nitrogen base (without amino acids, Sigma) supplemented with Synthetic Complete drop-out medium without uracil (Formedium) and 0.1% glucose. Protein expression was induced with 2% galactose for 40 h at 25 °C at an OD[600] of 0.8. For generation of selenomethionine labelled protein, BY4741 cells (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) were grown at 30 °C to an OD[600] of 2–3, centrifuged and washed to remove residual methionine before induction. The cells were subsequently suspended in yeast nitrogen base without amino acids (Sigma), supplemented with Synthetic Complete drop-out medium without Met/uracil (Formedium), 0.01% raffinose and 100 mg l⁻¹ Selenomethionine (Acros Organics), grown for 1 h, induced and expressed as described for wild type (WT). For expression in mammalian cells, tsA201 cells (catalogue no. 96121229, Sigma-Aldrich) with a confluency of 40–60% were transfected with plasmid DNA containing nhTMEM16 or mTMEM16A as described[50], except that the transfection buffer was prepared with 2.8 mM Na$_2$HPO$_4$. Expression was carried out in 10-cm dishes (Corning) at 37 °C and 2.2% CO$_2$ for 1–2 days. For electrophysiology HEK293T cells were transfected with the respective plasmids containing WT or mutant mTMEM16A (isoform ac, 5 μg of DNA per 3.5-cm dish) by similar protocols.

Protein purification. *S. cerevisiae* expressing WT nhTMEM16 was harvested by centrifugation and resuspended in buffer A (50 mM HEPES pH 7.6, 150 mM NaCl) containing 0.5 mM CaCl$_2$, protease inhibitors (Complete, Roche), DNase I, and 1 mM MgCl$_2$ and lysed in a custom-made pressure-based cell disruptor at 40,000 p.s.i. Cell debris was removed by low-spin centrifugation. Membranes were harvested by ultracentrifugation with a 45 TI rotor (Beckmann) at 40,000 r.p.m. for 1.5 h. All steps were carried out on ice or at 4 °C. Protein was extracted in buffer A containing 0.5 mM CaCl$_2$, 1% *n*-dodecyl-β-D-maltopyranoside (DDM, Anatrace) and protease inhibitors (Roche) for 1.5 h. Insoluble parts were removed by centrifugation for 30 min at 40,000 r.p.m. with a 45 TI rotor (Beckmann). After addition of 15 mM imidazole the protein was bound in batch to NiNTA for 1.5 h, washed with buffer B (10 mM HEPES pH 7.6, 150 mM NaCl, 5% glycerol, 0.025% DDM) containing 5 mM CaCl$_2$ and 50 mM imidazole and eluted in buffer B containing 5 mM CaCl$_2$ and 400 mM imidazole. The eluted fraction was cleaved with HRV 3C protease for 2 h and dialysed against buffer B containing 5 mM CaCl$_2$. The GFP–His[10] fragment was removed by binding to NiNTA resin, the flow-through was concentrated (Amicon) and applied to a Superdex 200 column (GE healthcare) equilibrated in buffer C (5 mM HEPES pH 7.6, 150 mM NaCl, 0.025% DDM) containing 3 mM CaCl$_2$. The peak fraction was concentrated to 8–14 mg ml⁻¹. Prior to crystallization 0.2% *n*-undecyl-α-D-maltopyranoside (Anatrace), 50 μg ml⁻¹ yeast polar lipid extract (solubilized in 1% DDM, Avanti Polar Lipids) and 2% 1,2,3-heptanetriol were added to the protein. The addition of additives was essential to remove the anisotropy of diffraction and improve the resolution from 6 to 3.3 Å. A 35 l fermentation culture harvested at an OD[600] of 4.5 typically yielded about 5 mg of pure protein. Details concerning the purification and crystallization of nhTMEM16 in Ca^{2+}-free conditions are described in the Supplementary Discussion.

For reconstitution into liposomes WT nhTMEM16 and the triple mutant containing an SBP tag were either purified from *S. cerevisiae* or HEK tsA201 cells with similar results. mTMEM16A was expressed in HEK tsA201 cells by the same protocol. HEK tsA201 cells or membranes of *S. cerevisiae* expressing the respective protein were collected and treated with buffer A containing 5 mM EDTA, 5% glycerol, protease inhibitors and 2% DDM. Cell debris was removed by centrifugation.

The supernatant was incubated with streptavidin resin (Pierce Streptavidin plus UltraLink) for 1.5 h and washed with buffer B. Protein was eluted with buffer B containing 2 mM biotin. The purity of the protein was confirmed by SDS–PAGE (Extended Data Fig. 2c). Initially, the protein was cleaved to remove the purification tag and subjected to size-exclusion chromatography on a Superdex 200 column before reconstitution. In later stages, the protein was reconstituted after affinity purification at 1 mg ml⁻¹ with very similar results. For reconstitution, an 18 l fermentation culture of *S. cerevisiae* typically yielded 400 μg of pure protein. All buffers used during reconstitution were made with Ca^{2+}-free water (Merck Milipore) and chemicals extra low in Ca^{2+}. Multi-angle light scattering (MALS) experiments were carried out at 20 °C on a HPLC (Agilent 1100) connected to an Eclipse 3 system equipped with a miniDAWN TREOS MALS detector and an Optilab T-rEX refractometer (Wyatt Technology). 50 μg of purified nhTMEM16 (1 mg ml⁻¹) were injected onto a Superdex S200 column equilibrated in buffer B and eluted protein was detected online. The molecular weight was calculated at each time point during elution using a combination of ultraviolet absorbance, light scattering and differential refractive index measurements with the Astra software package (Astra 6.0, Wyatt Technology). The determined molecular weight of the protein of about 145 kDa compares well with the predicted 166 kDa of the dimer.

Crystallization and structure determination. WT nhTMEM16 (containing two additional residues on the N terminus remaining from the protease cleavage site) was crystallized in sitting drops at 4 °C. Crystals were prepared by mixing protein at a concentration of 8–14 mg ml⁻¹ in a 1:1 ratio with reservoir containing either 100 mM Capso pH 9.4, 100 mM MgCl$_2$, 100 mM NaCl and 21–23% PEG400 (CF1) or 50 mM HEPES pH 7.4, 100 mM ammonium sulphate, 21–23% PEG400 (CF2). Crystals were harvested after 2–3 weeks (CF1) or 1 week (CF2), cryoprotected by increasing the PEG400 concentration to 36% and flash-frozen in liquid propane.

All data sets were collected on frozen crystals on the X06SA beamline at the Swiss Light Source (SLS) of the Paul Scherrer Institut (PSI) on a PILATUS 6M detector (Dectris, Extended Data Fig. 4a). The data were indexed, integrated and scaled with XDS[52] and further processed with CCP4 programs[53]. Both crystal forms are of space group P2$_1$2$_1$2$_1$ and each contains a dimer of the protein in its respective asymmetric unit (Extended Data Fig. 4a). The structure of the nhTMEM16 (CF2) was determined by the single-wavelength anomalous dispersion (SAD) method with data collected from crystals containing selenomethionine-derivatized protein. The Se-sites were identified with SHELX C and D[54,55] and refined in SHARP[56]. Initial phases at low resolution were improved by solvent flattening and twofold NCS averaging in DM[57]. A coarse model was built in O[58] and used as search model for molecular replacement in CF1 with PHASER[59]. Phases were subsequently extended to 3.3 Å by NCS and cross-crystal averaging with DM. Models were built with O and COOT[60]. The correct register of the protein was confirmed with the help of 13 methionine positions defined in the SeMet data set and from sulphur anomalous data collected for mutants F612M and L624M where methionine residues were inserted in regions of the protein that lack this amino acid. The structure was initially refined maintaining strict twofold NCS constraints in CNS[61]. In later stages, the strict constraints were loosened and restraint individual B-factors and TLS parameters were refined in PHENIX[62]. R and R$_{free}$ were monitored throughout. R$_{free}$ was calculated by selecting 5% of the reflection data that were omitted in refinement. The final model (CF1) contains 654 out of 735 residues per subunit, has R/R$_{free}$ values of 23.8% and 28.5%, good geometry and no outliers in the Ramachandran plot (Extended Data Fig. 4a). Regions not defined in the electron density include residues 1–18, 130–140, 465–482, 586–593, 657–659, 685–691 and 720–735. The structure of nhTMEM16 in CF2 was refined in PHENIX as described for CF1. Both structures show very similar conformations. Ca^{2+} positions were identified from data collected at 1.95 Å to improve the anomalous scattering of the bound ions and included in the refinement (Extended Data Fig. 4a).

Liposome preparation and scrambling assay. Liposomes were prepared as 3:1 mixture of *Escherichia coli* polar lipids/egg PC (Avanti Polar Lipids). For scramblase assays lipids were supplemented with either 0.5% 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(NBD) or 1,2-dioleoyl-*sn*-glycero-3-phospho-L-serine-*N*-(NBD) (Avanti Polar Lipids). For control experiments approximately 20 μM NBD-dextran (prepared following manufacturers instruction, Life technologies) was added during liposome preparation instead of NBD-labelled lipids. Liposomes were suspended in buffer D (20 mM HEPES pH 7.4, 300 mM KCl) containing either 2 mM EGTA (for Ca^{2+}-free conditions), or 2 mM EGTA and the concentration of Ca^{2+} or other divalent cations (made from the respective nitrate salts) as calculated by MAXCHELATOR (http://maxchelator.stanford.edu/CaMgATPEGTA-TS.htm) to reach the indicated free divalent ion concentration. Liposomes were prepared as described[63]. Briefly, liposomes were subjected to three freeze-thaw cycles, subsequently extruded through a 400-nm polycarbonate filter (Avestin) and destabilized with Triton-X-100. Purified protein (5 μg per mg lipid) was added and detergent was removed by stepwise addition of SM-2 adsorbent biobeads (Bio-Rad). Proteoliposomes were formed at 4 °C under gentle agitation, incubated for 40 h, collected

by ultracentrifugation, resuspended in buffer D containing the above-mentioned concentrations of EGTA and divalent ions at a lipid concentration of 20 mg ml^{-1}, flash-frozen in liquid N_2 and stored at $-80\,°C$. All buffers were prepared with Ca^{2+}-free water (Merck Millipore) using highly pure chemicals low in Ca^{2+}. The scramblase assay was performed similarly as previously described[14]. After three freeze-thaw cycles and extrusion (400-nm filter), 20 μl of proteoliposome suspension was diluted in 2 ml buffer D (with HEPES pH 7.4 concentration increased to 60 mM) containing either 2 mM EGTA, or 2 mM EGTA and the calculated concentrations of divalent ions, in a stirred cuvette at 23 °C. Sodium dithionite (Sigma) was added after 1 min to a final concentration of 30 mM unless stated otherwise and fluorescence decay was recorded on a Fluoromax-4 spectrofluorometer (Horiba, excitation 470 nm, emission 530 nm). For analysis the fluorescence intensity was normalized to F/F_{max}.

Patch-clamp electrophysiology. For electrophysiology mTMEM16A–YFP and nhTMEM16–YFP were expressed in HEK293T cells. Cells expressing either protein were identified by the fluorescence of the C-terminal Venus–YFP tag and used for patch-clamp experiments within 36 h after transfection. Experiments were conducted at room temperature (20–22 °C) with fire-polished borosilicate glass patch pipettes (4–8 MΩ). Currents were recorded in either whole-cell configuration or from excised patches in the inside-out configuration with an Axopatch 200B amplifier, digitized at 10 kHz, filtered at 1 kHz and analysed using Clampfit (MDS Analytical Technologies). Solutions were prepared as described[16]. Standard external solution contained: 140 mM NaCl, 4 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM glucose, and 10 mM HEPES (pH 7.4). Ca^{2+}-free intracellular slution contained 146 mM $CsCl_2$, 2 mM $MgCl_2$, 5 mM EGTA, 10 mM sucrose, and 8 mM HEPES (pH 7.4), adjusted with N-methyl-D-glucamine. High Ca^{2+} solution contained 5 mM Ca^{2+}-EGTA (resulting in a free Ca^{2+} concentration of 20 μM). Intermediate Ca^{2+} solutions were prepared by mixing Ca^{2+}-free and high-Ca^{2+} solutions in corresponding ratios. Solution containing free Ca^{2+} concentrations higher than 20 μM were prepared by addition of the corresponding amounts of $CaCl_2$. In these cases EGTA was replaced by Br_2-BAPTA (5,5′-dibromo-1,2-bis(2-aminophenoxy)ethane-N,N,N′,N′-tetraacetic acid, 3.5 mM; Invitrogen). Solutions were applied with a double-barrelled theta tubing with a tip diameter of 400 μm attached to a piezo-bimorph (Siskiyou). The Ca^{2+} dependence of mTMEM16A and its mutants was measured in excised inside-out patches. Whereas the observed maximum current was similar to WT in mutants N650A, E702Q and E705Q, it was generally smaller in mutants E734Q, D738N and E654A. Currents generally saturated at high Ca^{2+} concentration except for E654A and D738N, where they continue to increase even at concentrations up to 10 mM Ca^{2+}. Activation in D738N shows a biphasic behaviour with an apparent saturation of currents around 100 μM and a subsequent increase above 500 μM Ca^{2+}. (Extended Data Fig. 9 and 10). A decrease of the response at increasing Ca^{2+} concentrations indicates rundown. For analysis of dose–response relationships, the current-response at different Ca^{2+} concentration recorded at a holding potential of 80 mV, was fitted to a Hill equation. Responses in D738N were only considered up to a Ca^{2+} concentration of 500 μM. The averages in the EC_{50} from 3–4 independent recordings (Extended Data Fig. 9 and 10) show shifts in the EC_{50} towards higher Ca^{2+} concentrations for all investigated mutants (WT, EC_{50} 0.36 μM, n 2.5; N650A, EC_{50} 1.8 μM, n 2.6; E702Q, EC_{50} 9.5 μM, n 2.1; E705Q, EC_{50} 231 μM, n 1.0; E734Q, EC_{50} 4.0 μM, n 1.6; D738N, EC_{50} 20.0 μM, n 0.84, where n is the Hill coefficient). To demonstrate the statistical significance of this increase, EC_{50} values were log-transformed for one-way ANOVA and subsequently compared to WT values using Tukey's post-hoc test for significance. Values were considered significantly different if $P < 0.05$. The analysis revealed that all shifts in the EC_{50} of Ca^{2+}-binding site mutants are statistically significant. The voltage dependence of WT suggests that Ca^{2+} crosses about 18% of the transmembrane electric field to reach its binding site (Fig. 5b).

Planar lipid bilayer experiments. For recording in planar lipid bilayers, nhTMEM16 was purified and reconstituted similarly as for the scramblase assay

at lipid to protein ratios of 100:1 or 200:1, except that no NBD lipids were added. The incorporation of the protein into liposome was confirmed by freeze-fracture electron microscopy as described[64]. Proteoliposomes containing nhTMEM16 were fused to bilayers formed from 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1′-rac-glycerol) (in ratio of 1:3 w/w, Avanti) and recorded with a horizontal planar lipid bilayer system as described[65,66]. In recordings under symmetric ion concentrations, both chambers contained 10 mM HEPES pH 7.4, 150 mM NaCl (buffer) and either no or 300 μM $CaCl_2$. In recordings under asymmetric conditions the NaCl concentration in one chamber was reduced to 15 mM. Electrodes were connected to the respective bath solutions via salt bridges. Currents were recorded with an Axopatch 200B amplifier, digitized at 10 kHz, filtered at 1 kHz and analysed using Clampfit.

47. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50,** 3272–3278 (2011).
48. Keefe, A. D., Wilson, D. S., Seelig, B. & Szostak, J. W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-tag. *Protein Expr. Purif.* **23,** 440–446 (2001).
49. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnol.* **20,** 87–90 (2002).
50. Schenck, S., Wojcik, S. M., Brose, N. & Takamori, S. A chloride conductance in VGLUT1 underlies maximal glutamate loading into synaptic vesicles. *Nature Neurosci.* **12,** 156–162 (2009).
51. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature Protocols* **3,** 784–798 (2008).
52. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
53. The Collaborative Computational Project, Number 4. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
54. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
55. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37,** 843–844 (2004).
56. de La Fortelle, E. & Bricogne, G. in *Methods in Enzymology* Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 492–494 (Academic, 1997).
57. Cowtan, K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38 (1994).
58. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
59. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
61. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
62. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
63. Geertsma, E. R., Nik Mahmood, N. A., Schuurman-Wolters, G. K. & Poolman, B. Membrane reconstitution of ABC transporters and assays of translocator function. *Nature Protocols* **3,** 256–266 (2008).
64. Garcia-Celma, J., Szydelko, A. & Dutzler, R. Functional characterization of a ClC transporter by solid-supported membrane electrophysiology. *J. Gen. Physiol.* **141,** 479–491 (2013).
65. Accardi, A., Kolmakova-Partensky, L., Williams, C. & Miller, C. Ionic currents mediated by a prokaryotic homologue of ClC Cl$^-$ channels. *J. Gen. Physiol.* **123,** 109–119 (2004).
66. Zimmermann, I. & Dutzler, R. Ligand activation of the prokaryotic pentameric ligand-gated ion channel ELIC. *PLoS Biol.* **9,** e1001101 (2011).
67. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).

a
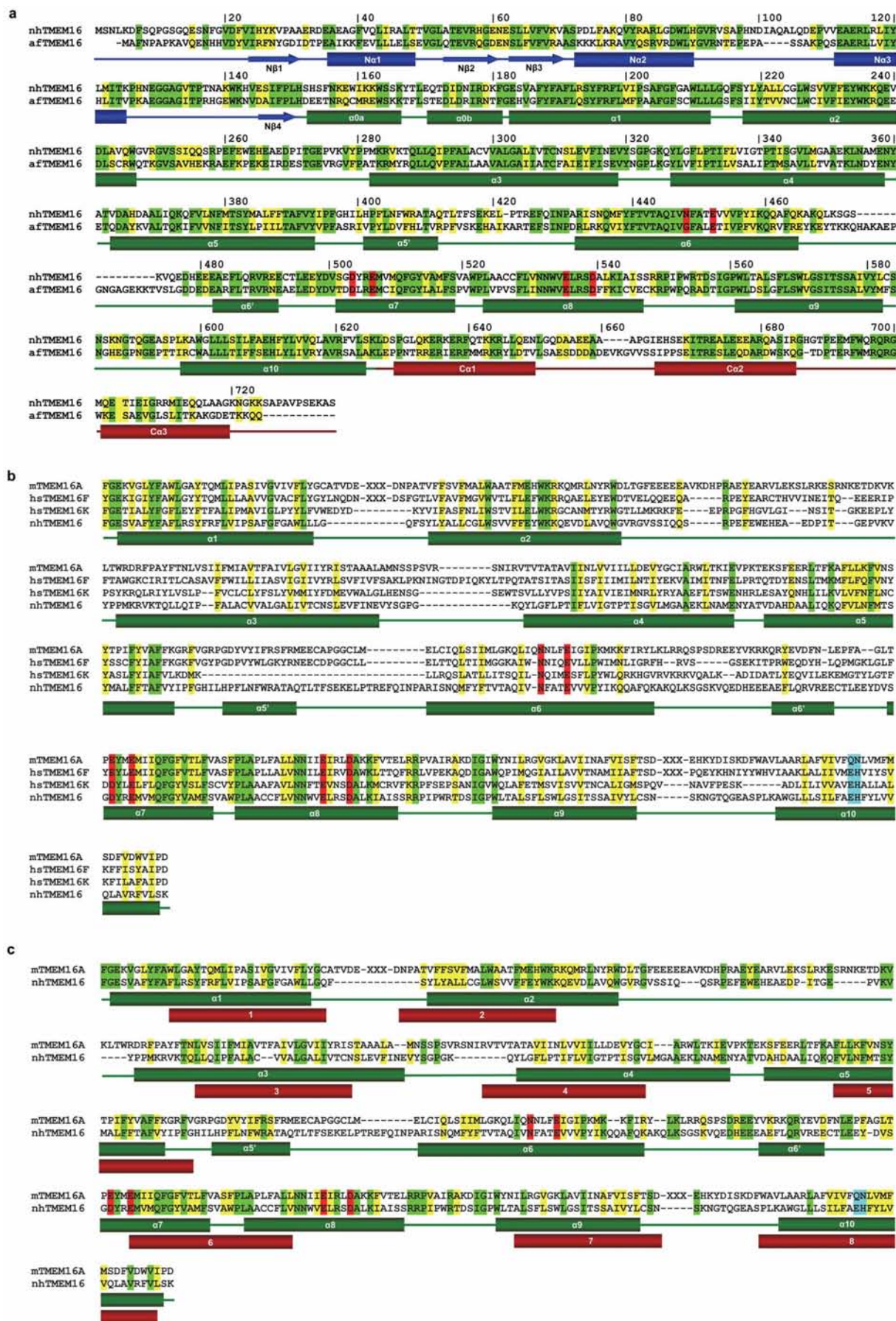
Extended Data Figure 1 | Structure-based sequence alignment. Sequences were aligned with Clustal Omega[67] and edited manually. Identical residues are highlighted in green, homologous residues in yellow and residues of the Ca^{2+}-binding site in red. Secondary structure elements are shown below. **a**, Comparison of nhTMEM16 and afTMEM16. The numbering corresponds to nhTMEM16. **b**, Comparison of the membrane domains of selected TMEM16 proteins, m refers to murine, hs to human proteins. Long insertions in loop regions of mammalian family members (indicated by -xxx-) are not shown in the alignment. The positions of residues in α-helix 10 involved in an inter-subunit salt bridge at the dimer interface are highlighted in cyan. **c**, Comparison of the observed and predicted topology of TMEM16 proteins. Sequence alignment of the membrane spanning regions of mTMEM16A and nhTMEM16 with the observed (green) and predicted topology[25] (red) indicated. Identical residues are highlighted in green, homologous residues in yellow, residues of the Ca^{2+}-binding site in red and the inter-subunit salt bridge at the dimer interface in cyan. The difference between the predicted and observed transmembrane segments is due to the failure of sequence-based approaches to identify the correct boundaries of several helices and to detect α-helix 6 at all and helices 7 and 8 as separate entities.
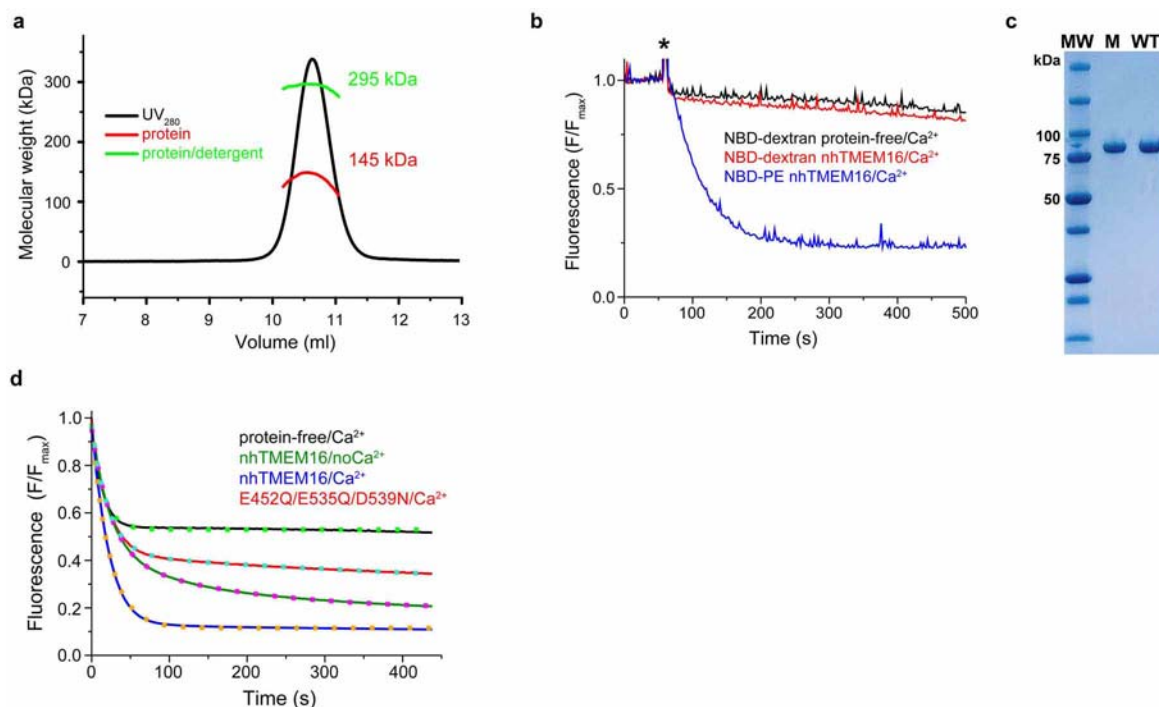
Extended Data Figure 2 | Multi-angle light scattering and lipid scrambling.
a, Gel filtration and light scattering results for nhTMEM16 in the detergent DDM. The continuous black trace corresponds to the absorption at 280 nm. Molecular weights of the protein and the protein-detergent complex are shown in red and green, respectively. **b**, Inaccessibility of NBD groups trapped within liposomes. Dithionite is incapable of reducing the soluble NBD-dextran trapped in the interior of proteoliposomes containing nhTMEM16. Traces of proteoliposomes containing nhTMEM16 and empty liposomes are shown in red and black respectively. Traces from proteoliposomes of nhTMEM16 containing NBD-PE at equivalent dithionite levels are shown for comparison (blue). Asterisk marks addition of 2.5 mM dithionite. **c**, SDS–PAGE gel of the Ca^{2+}-binding site triple-mutant E452Q/E535Q/D539N (M) and nhTMEM16 (WT) used for reconstitution illustrating the purity of the sample. The molecular weight marker (MW) is shown on the left with selected bands labelled. **d**, Analysis of phospholipid scrambling. Time dependent fluorescence decrease of NBD-PE upon reduction by 30 mM dithionite ($t = 0$). The traces are as in Figs 1b–d and 5a. A fit to a single exponential decay is shown as dotted lines for protein-free/Ca^{2+} and nhTMEM16/Ca^{2+} with time constants of 15 s and 22 s, respectively. A fit to a sum of two exponential functions is shown for nhTMEM16/no Ca^{2+} and E452Q/E535Q/D539N/Ca^{2+} with time constants of 25 and 21 s for the fast component and 175 and 803 s for the slow component, respectively.
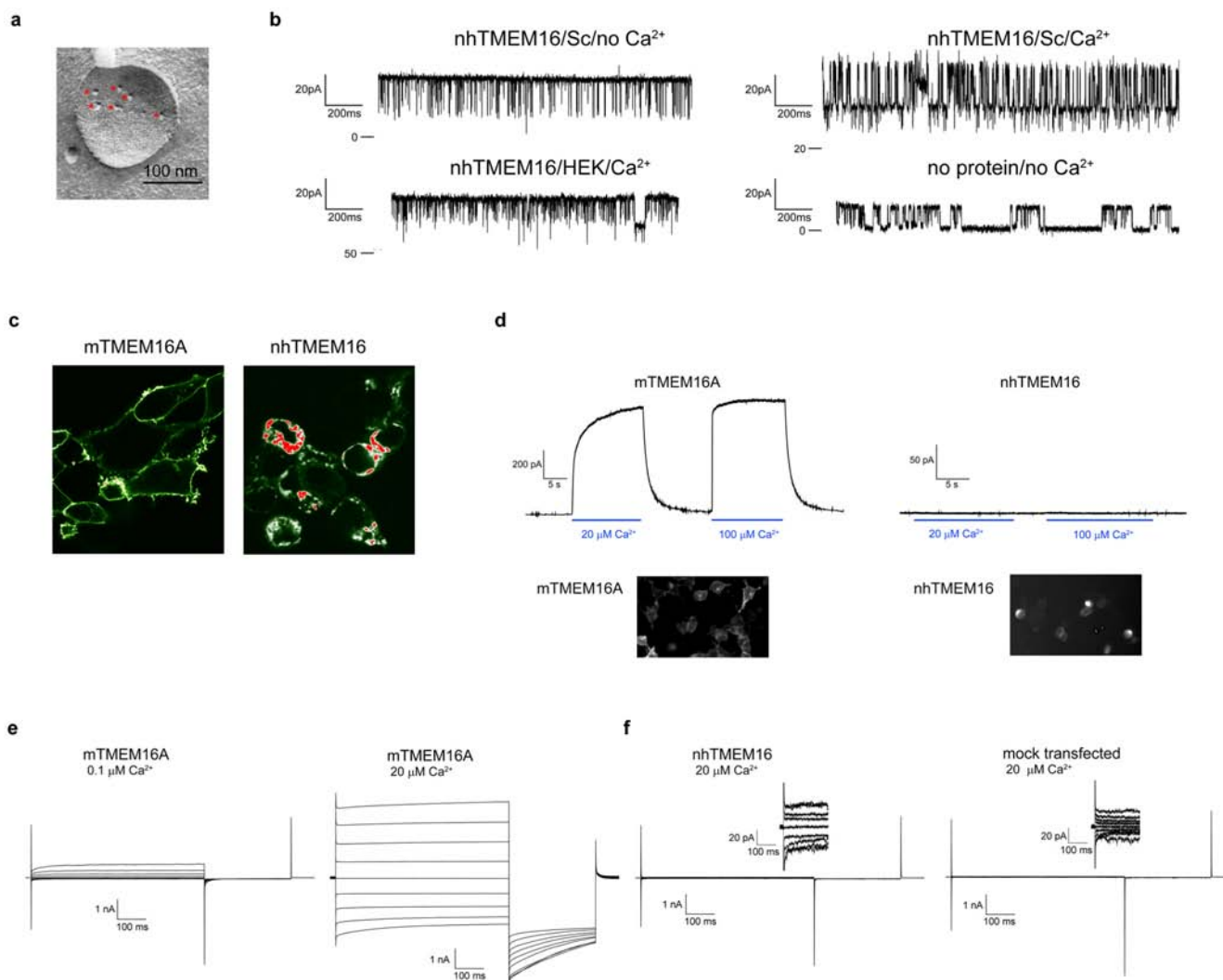
Extended Data Figure 3 | Search for ion channel activity in nhTMEM16.
a, Freeze-fracture electron microscopy image of a proteoliposome containing nhTMEM16 formed from a 3:1 mixture of *E. coli* polar lipids/egg PC. Reconstituted proteins are labelled with red asterisks. **b**, Planar lipid bilayer experiments. Currents recorded after fusion of proteoliposomes containing nhTMEM16 expressed and purified from *S. cerevisiae* (Sc) in the absence of Ca^{2+} (top, left), with 300 μM Ca^{2+} added on both sides of the bilayer (top right) and of proteoliposomes containing nhTMEM16 expressed and purified from HEK tsA201 cells in the presence of 300 μM Ca^{2+} added on both sides of the bilayer (bottom left). Currents recorded after fusion of liposomes of the same lipid composition not containing any protein are shown for comparison (bottom right). Displayed traces were recorded at a holding potential of 100 mV in symmetric solutions containing 150 mM NaCl and 10 mM HEPES pH 7.4. Selected current levels (in pA) are indicated on the left. **c**, Fluorescence confocal microscopy images of HEK tsA201 cells expressing a mTMEM16A–YFP fusion construct (left) or a nhTMEM16–YFP fusion construct (right). **d**, Recordings from excised inside-out patches.

Representative current response in a membrane patch excised from cells expressing a mTMEM16A–YFP fusion construct upon rapid exchange into solutions containing the indicated amount of Ca^{2+} (left) and equivalent recordings from patches that were excised from cells expressing a nhTMEM16–YFP fusion protein (right). The voltage was clamped at 80 mV. The fluorescence of transfected cells expressing mTMEM16A–YFP used for recording is shown below. No activity of nhTMEM16–YFP was observed in any of more than 30 patches. **e**, Patch-clamp recording in the whole-cell configuration. Representative currents from a HEK293T cell expressing a mTMEM16A–YFP construct recorded from a solution containing either 0.1 μM (left) or 20 μM (right) free Ca^{2+} in the patch pipette. **f**, Representative currents from a cell expressing a nhTMEM16–YFP fusion protein recorded from a solution containing 20 μM free Ca^{2+} in the patch pipette (left). Current response from mock-transfected cells recorded under the same conditions is shown for comparison (right). Insets show part of the traces with magnified current scale.

a

Table Data collection and refinement statistics

	nhTMEM16/CF1	nhTMEM16/CF2	nhTMEM16/SeMet	nhTMEM16/anom	nhTMEM16/no Ca²⁺
Data collection					
Wavelength (Å)	0.9797	1.0	0.9797	1.95	1.95
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, b, c (Å)	96.5, 113.7, 235.7	115.9, 127.2, 180.1	113.7, 124.8, 177.4	115.2, 124.8, 177.4	115.2, 127.1, 179.7
(°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	50-3.3 (3.4-3.3)*	50-3.4 (3.5-3.4)	50-4.0 (4.1-4.0)	50-3.5 (3.6-3.5)	50-4.2 (4.3-4.2)
R_{merge}	8.3 (123.5)	6.5 (149.3)	11.7 (137.9)	10.2 (116.5)	15.2 (233.7)
$I/\sigma I$	20.1 (2.6)	20.0 (1.7)	18.4 (2.3)	16.5 (1.8)	12.5 (2.0)
Completeness (%)	99.1 (98.8)	98.9 (87.9)	99.0 (100)	99.9 (100.0)	99.9 (100)
Redundancy	12.7 (12.7)	9.6 (8.6)	22.9 (15.9)	12.2 (7.0)	18.8 (17.8)
$CC_{1/2}$ (%)	99.9 (80.3)	99.9 (59.8)	100.0 (71.1)	99.8 (69.5)	99.8 (75.3)
Refinement					
Resolution (Å)	15-3.3	15-3.4		15-3.5	15-4.2
No. reflections	38985	36750		32709	19356
R_{work}/R_{free}	23.8 (28.5)	24.8 (29.2)		23.7 (28.5)	23.0 (27.2)
No. atoms	10574	10574		10574	10570
Protein	10570	10570		10570	10570
Ligand/ion	4	4		4	0
B-factors					
Protein	137	159		147	199
Ligand/ion	104	146		123	-
R.m.s deviations					
Bond lengths (Å)	0.003	0.002		0.003	0.003
Bond angles (°)	0.74	0.70		0.90	0.78

*Highest resolution shell is shown in parenthesis.

b
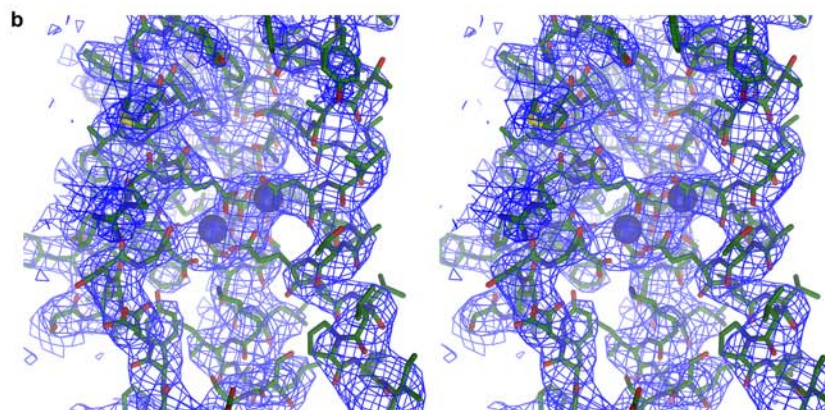
Extended Data Figure 4 | Crystallography. a, Table describing data collection and refinement statistics of five data sets presented in this study. nhTMEM16/CF1 and nhTMEM16/CF2 are data sets used for the building and refinement of the crystal structures of CF1 and CF2 respectively that have been deposited in the PDB. nhTMEM16/SeMet, a data set of a selenomethionine derivative collected at the Se anomalous absorption edge, was used for obtaining initial phases of CF2. nhTMEM16/anom is a data set used for the identification of the Ca²⁺-binding site by anomalous scattering and nhTMEM16/no Ca²⁺ is from a protein purified in the presence of EDTA and crystallized without addition of Ca²⁺. **b**, Stereo view of the Ca²⁺-binding region in CF1. The model of the protein displayed as sticks is shown with experimental electron density superimposed. The map was calculated at 3.3 Å with Se-Met SAD phases that were improved by solvent flattening, cyclic twofold NCS and cross-crystal averaging (blue mesh, contoured at 1σ). Ca²⁺ ions are shown as blue spheres.
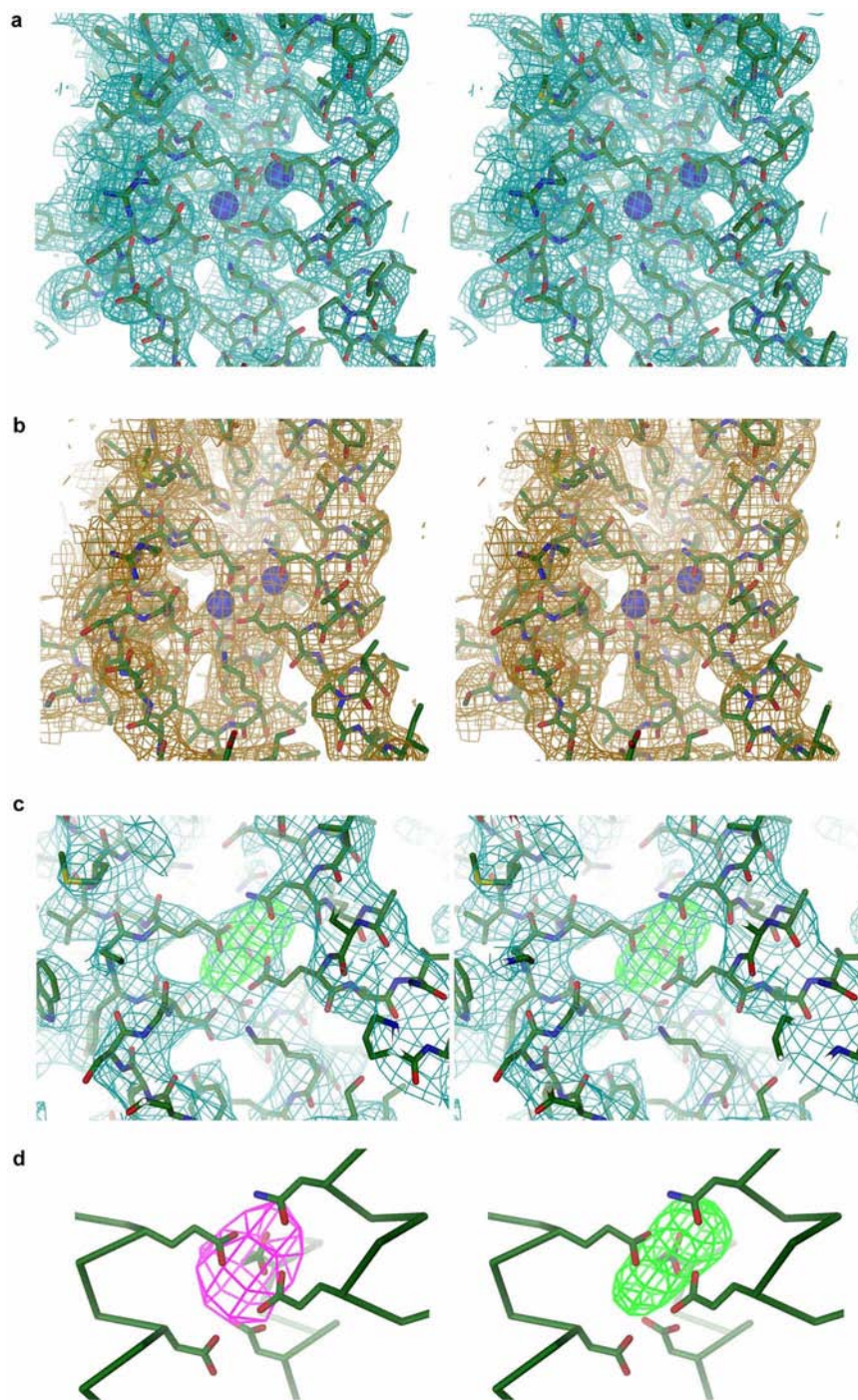
Extended Data Figure 5 | Electron density. a, Stereo view of the Ca^{2+}-binding region in CF1. The model of the protein displayed as sticks is shown with $2F_o - F_c$ electron density superimposed (cyan mesh, contoured at 1σ after sharpening with $b = 50$). The density at 3.3 Å was calculated with phases from the refined model. Ca^{2+} ions are shown as blue spheres. **b**, $2F_o - F_c$ electron density of the Ca^{2+}-binding region in CF2 (calculated at 3.4 Å and contoured at 1σ after sharpening with $b = 50$, orange) superimposed on the refined model. **c**, Stereo view of the Ca^{2+}-binding region of a structure obtained from protein purified in the presence of EDTA and crystallized in CF2 without addition of Ca^{2+}. $2F_o - F_c$ electron density (cyan mesh, calculated at 4.2 Å and contoured at 1σ after sharpening with a $b = 50$) and $F_o - F_c$ density (contoured at 3σ, green) is superimposed on the refined model. No ions were included in the refinement. **d**, Close-up of the Ca^{2+}-binding site. Anomalous difference density (left, calculated at 6 Å and contoured at 4σ, magenta) and $F_o - F_c$ density (right, contoured at 3σ, green) indicates the presence of bound Ca^{2+} ions.
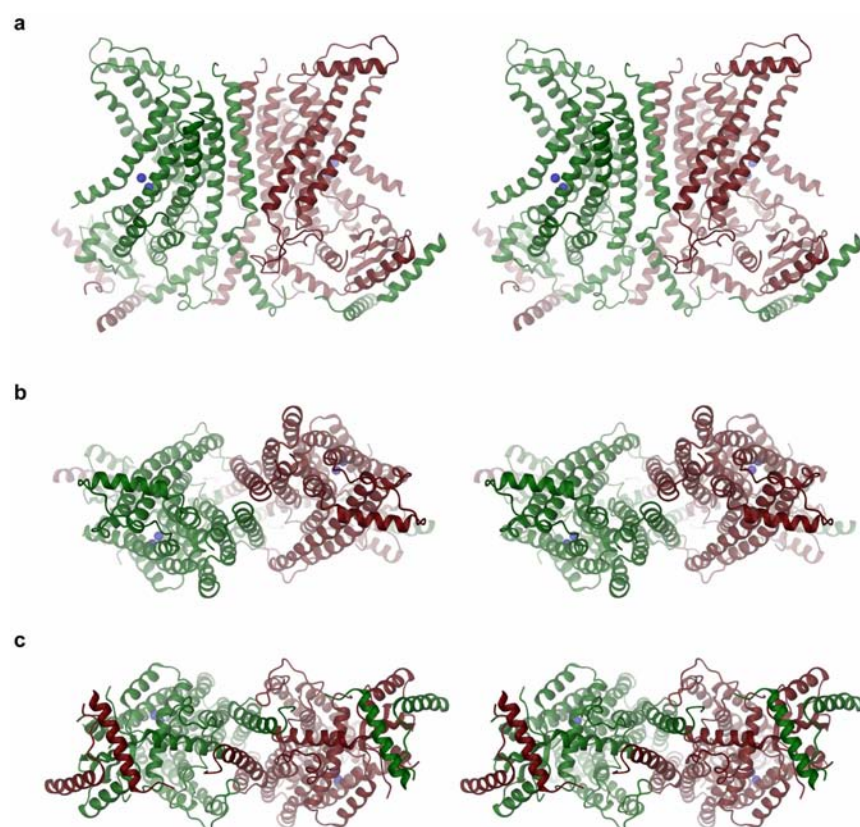
Extended Data Figure 6 | nhTMEM16 dimer. Stereo views of a ribbon representation of the dimeric protein. Bound Ca^{2+} ions are shown as blue spheres. **a**, View from within the membrane; **b**, view from the extracellular side; **c**, view from the cytoplasm.
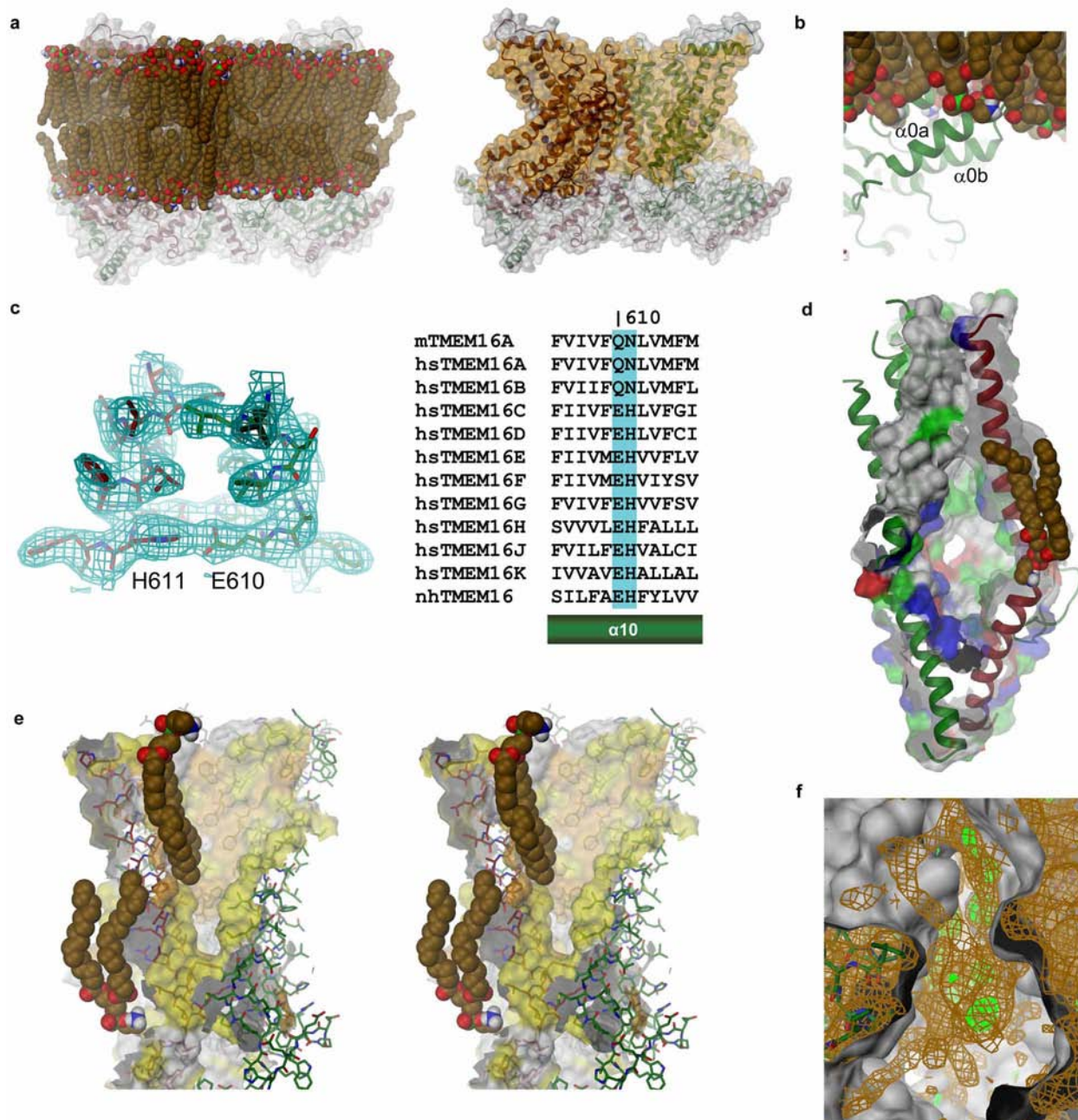
Extended Data Figure 7 | Model of lipid interactions and dimer cavity.
a, Model of nhTMEM16 embedded in a lipid membrane (left). The protein was positioned within the model of a PC bilayer (obtained from http://www.lobos.nih.gov/mbs/coords.shtml). A ribbon representation of the protein and the molecular surface are shown. Lipids are displayed as CPK models. Same view of the protein with regions on the surface presumably in contact with the membrane coloured in orange (right). **b**, Putative location of α-helices 0a and 0b relative to the lipid bilayer. **c**, Inter-subunit interactions between residues of α-helix 10. The protein is shown as sticks with $2F_o - F_c$ density (CF1, calculated at 3.3 Å and contoured at 1σ after sharpening with $b = 50$, cyan mesh) superimposed (left). A sequence alignment of the corresponding region underlines the conservation of interacting residues. Amino acids of the salt bridge in nhTMEM16 are highlighted in cyan, the numbering corresponds to nhTMEM16. **d**, View on the dimer cavity from the dimer interface. The molecular surface is coloured according to the properties of contacting residues (red, acidic; blue, basic; green, polar). A modelled lipid indicates the boundary of the inner leaflet of the bilayer. **e**, Stereo view of the cleft between α-helices 3 and 10. The protein is shown as stick model. The molecular surface is coloured according to the properties of contacting residues (yellow, hydrophobic; orange, aromatic). Lipids indicate the membrane boundary. **f**, Residual density in the dimer cavity. The molecular surface is coloured in white. $2F_o - F_c$ density (CF2, contoured at 1σ after sharpening with $b = 50$, orange) and $F_o - F_c$ density (contoured at 3σ, green) are shown. The view is as in **d**.
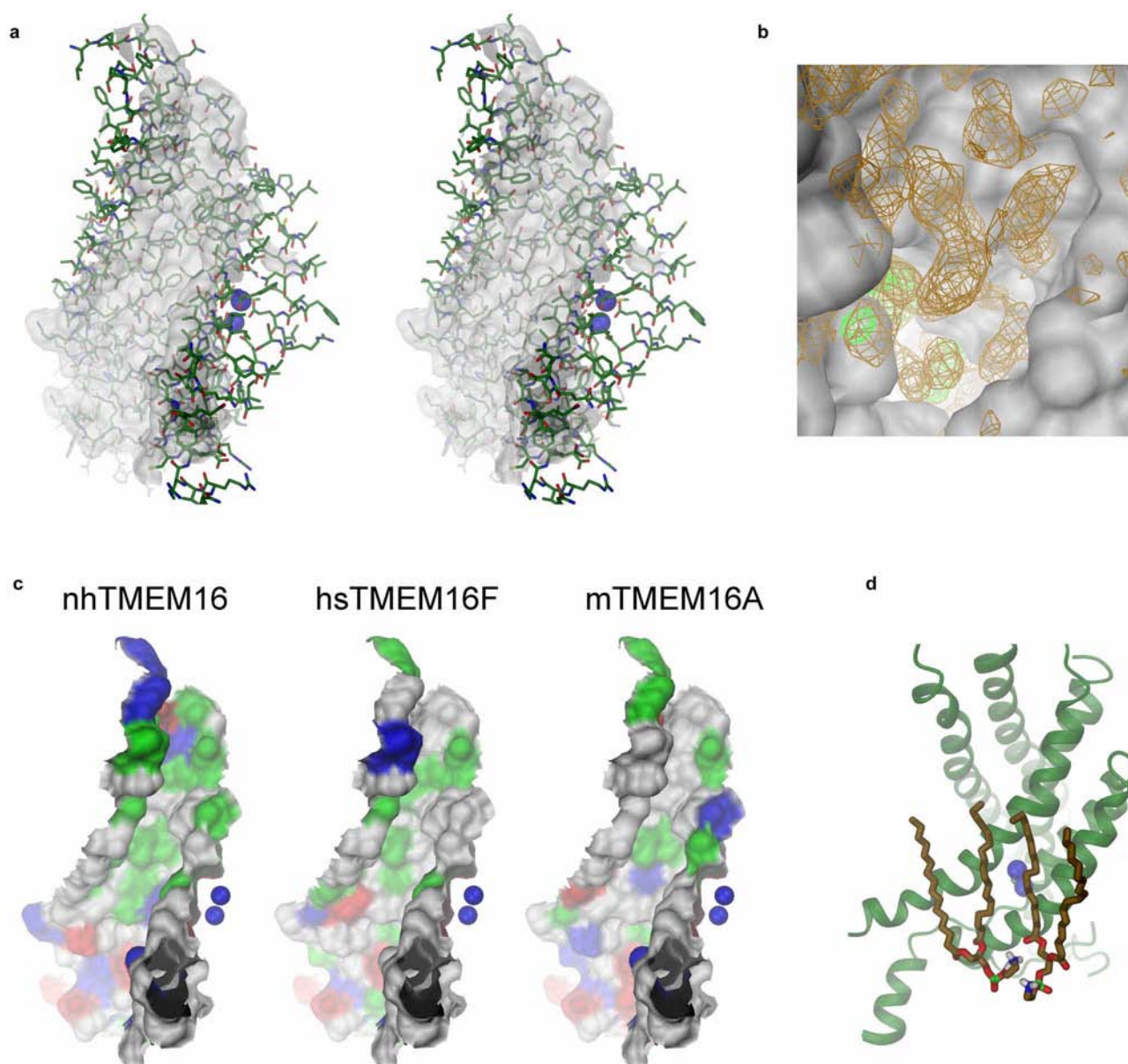
Extended Data Figure 8 | Subunit cavity and Ca²⁺-binding site. a, Stereo view of the subunit cavity viewed from within the membrane. Protein residues and the molecular surface are shown. **b,** Residual density in the subunit cavity. The molecular surface of the protein is shown. $2F_o - F_c$ density (CF2, contoured at 1σ after sharpening with $b = 50$, orange) and $F_o - F_c$ density (contoured at 3σ, green) are displayed. **c,** Model of the subunit cavity in different TMEM16 proteins. The molecular surface is coloured according to the properties of contacting residues (red, acidic; blue, basic; green, polar). Putative surface-exposed residues were obtained from a sequence alignment with nhTMEM16. **d,** Location of the Ca²⁺-binding site in relation to the lipid bilayer. Modelled lipids of the inner leaflet of the bilayer are shown as sticks.
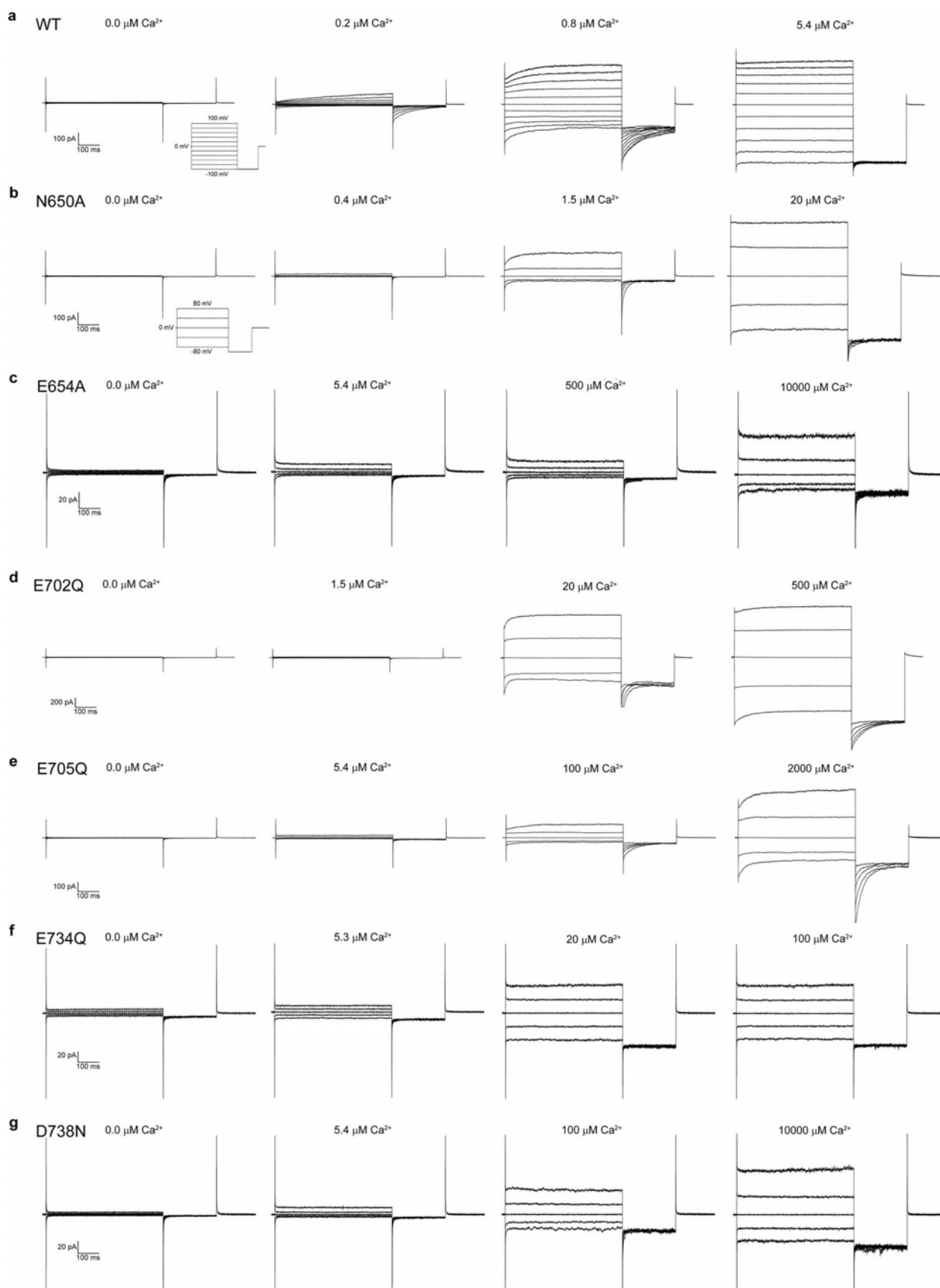
Extended Data Figure 9 | Electrophysiology. Current response in HEK293T cells overexpressing mTMEM16A–YFP and point mutants of the Ca^{2+}-binding site. All recordings were measured from single excised patches in the inside-out configuration after changing to intracellular solutions containing the indicated Ca^{2+} concentrations. **a**, WT, with voltage protocol shown as inset. **b**, Mutant N650A, with the voltage protocol shown as inset. **c–g**, Recordings of mutants E654A, **c**, E702Q, **d**, E705Q, **e**, E634Q, **f**, and mutant D738N, **g**.
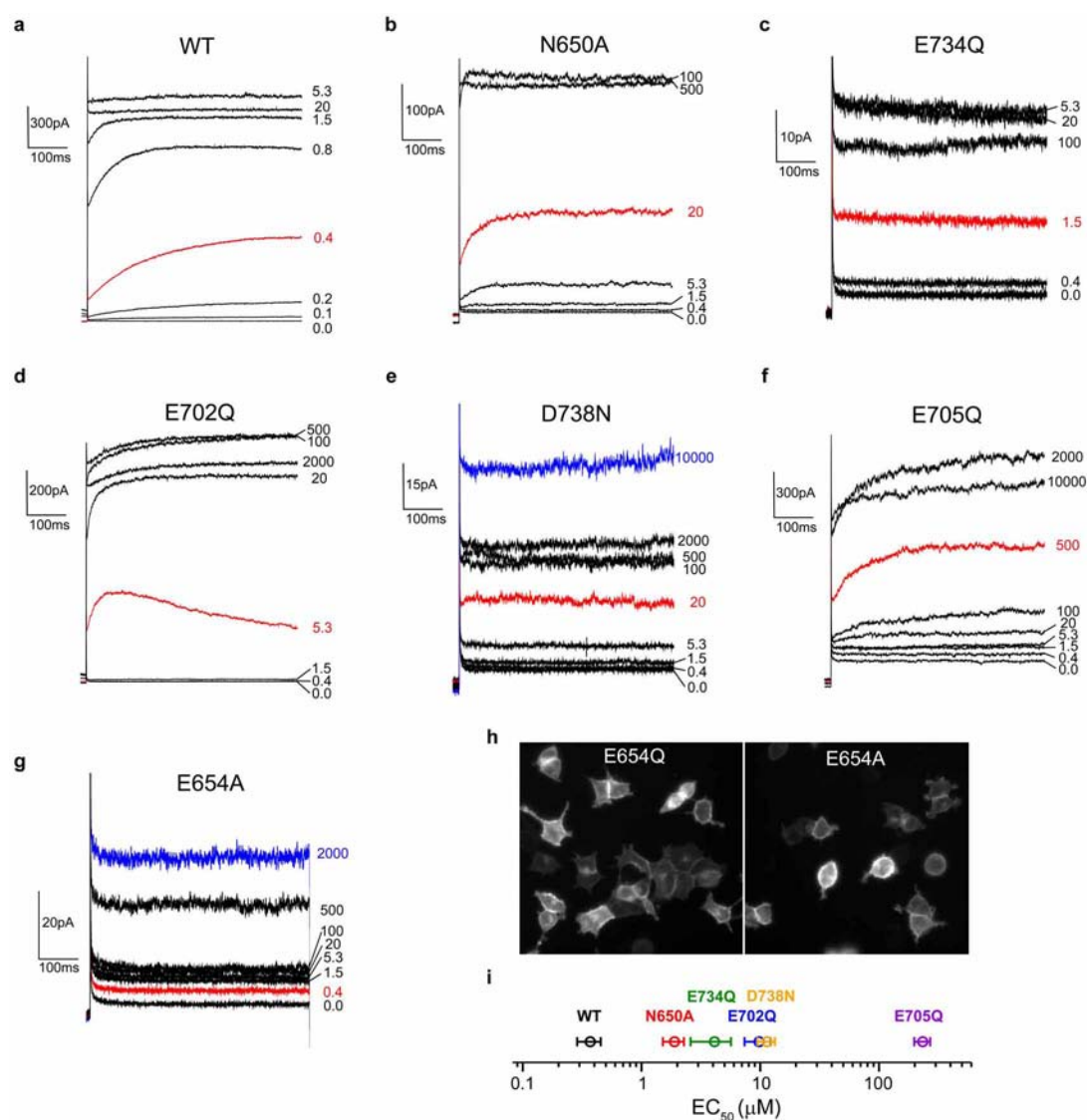
Extended Data Figure 10 | Ca²⁺ activation of mTMEM16A. Representative current traces of mTMEM16 and mutants of the Ca^{2+}-binding site. Currents were measured from excised inside-out patches of HEK293T cells expressing the respective protein at 80 mV. The Ca^{2+} concentration is indicated, selected traces are shown in colour. **a**, WT, **b**, N650A, **c**, E734Q, **d**, E702Q, **e**, D738N **f**, E705Q and **g**, E654A. **h**, Fluorescence microscopy image of mTMEM16A mutants expressed in HEK293T cells. Expression of fluorescently labelled protein is shown for mutants E654Q (left) and E654A (right). **i**, Analysis of the EC$_{50}$ of Ca^{2+} activation for different binding site mutants. The data show averages of fits to 3–4 independent recordings. Errors are s.d.

ARTICLE

Structure and insights into the function of a Ca^{2+}-activated Cl^- channel

Veronica Kane Dickson[1], Leanne Pedi[1] & Stephen B. Long[1]

Bestrophin calcium-activated chloride channels (CaCCs) regulate the flow of chloride and other monovalent anions across cellular membranes in response to intracellular calcium (Ca^{2+}) levels. Mutations in bestrophin 1 (BEST1) cause certain eye diseases. Here we present X-ray structures of chicken BEST1–Fab complexes, at 2.85 Å resolution, with permeant anions and Ca^{2+}. Representing, to our knowledge, the first structure of a CaCC, the eukaryotic BEST1 channel, which recapitulates CaCC function in liposomes, is formed from a pentameric assembly of subunits. Ca^{2+} binds to the channel's large cytosolic region. A single ion pore, approximately 95 Å in length, is located along the central axis and contains at least 15 binding sites for anions. A hydrophobic neck within the pore probably forms the gate. Phenylalanine residues within it may coordinate permeating anions via anion–π interactions. Conformational changes observed near the 'Ca²⁺ clasp' hint at the mechanism of Ca^{2+}-dependent gating. Disease-causing mutations are prevalent within the gating apparatus.

Ca^{2+}-activated Cl^- channels (CaCCs) are present in most eukaryotic cell types and are implicated in diverse functions including phototransduction, olfactory transduction, neuronal and cardiac excitability, smooth muscle contraction, and epithelial Cl^- secretion[1]. Bestrophin proteins constitute a family of CaCCs, distinct from the TMEM16 family[2–4], that open their anion-selective pores in response to a rise in the intracellular Ca^{2+} concentration[5–8]. Bestrophins have broad tissue distribution and, while their physiological roles are not fully known, evidence suggests that they function not only at the plasma membrane but also in other intracellular organelles[7,9].

Humans have four bestrophin paralogues (BEST1, BEST2, BEST3 and BEST4) that form CaCCs in the plasma membrane when expressed[5–7,10–12]. The highly conserved amino-terminal region of the proteins (amino acids 1–390; >55% sequence identity) is sufficient for CaCC activity[13]. The carboxy-terminal region (amino acids 391–585 of BEST1) has low sequence identity and is predicted to be unstructured. Approximately 200 mutations in BEST1 have been associated with retinal degenerative disorders, most commonly with vitelliform macular dystrophy (Best's disease), but also with other retinopathies[7,14–20]. Almost all of these occur within the N-terminal region. Although the steps leading to the disease state are not fully understood, most of the characterized mutations alter electrophysiological properties of the channel[5,13,16,21–24].

Bestrophin channels bear no discernable sequence homology with other ion channel families and no structural information is available for them. Properties including subunit topology and stoichiometry are unresolved. One recent study using the single-molecule photobleaching technique led the authors to conclude that bestrophins are tetramers[25], while other experiments suggest pentameric stoichiometry[5].

Partly because CaCC function has yet to be demonstrated using purified protein, there has been some debate about whether bestrophin is a channel or whether it is a modulator of other channels[7]. However, the effects of mutations (for example, see refs 11, 13) bolster the view that assembled bestrophin subunits contain Cl^--conducting pore(s) and that pore gating is regulated by direct binding of Ca^{2+} to a cytosolic region of the channel (K_d ~150 nM) that might involve a highly conserved cluster of acidic residues[5,6,12,26,27].

In addition to Cl^-, BEST1 conducts other monovalent anions including bromine (Br^-), iodine (I^-), thiocyanate (SCN^-), bicarbonate (HCO_3^-) and nitrate (NO_3^-)[7,28,29]. In contrast, the channel is essentially impermeable to the divalent sulphate anion (SO_4^{2-})[7,28]. Previous results suggest that mammalian BEST1 has permeability to GABA (γ-aminobutyric acid) and glutamate and that these permeabilities underlie a tonic form synaptic inhibition in the central nervous system and glutamate release from astrocytes, respectively[30,31].

For a better understanding of the architecture of bestrophin, its mechanisms for ion permeation, ion selectivity and Ca^{2+}-dependent gating, and the effects of disease-causing mutations, we have reconstituted CaCC function from purified protein and have determined X-ray structures of BEST1–Fab complexes with Ca^{2+} and permeant anions.

Crystallization of BEST1–Fab complexes

A construct encompassing amino acids 1–405 of chicken BEST1 ($BEST1_{cryst}$), which shares 74% sequence identity with human BEST1 (Extended Data Fig. 1), exhibited good biochemical stability and was selected for crystallization (Methods). Well-ordered crystals formed in the presence of trace amounts (~1 µm) of Ca^{2+} and required crystallization with a Fab monoclonal antibody fragment that preferentially recognizes the Ca^{2+}-bound form of $BEST1_{cryst}$ (Extended Data Fig. 2). Crystals obtained at pH 8.5 (space group C2) and at pH 4.0 (space group $P2_1$) diffracted X-rays to 3.1 Å and 2.85 Å resolution, respectively (Extended Data Table 1). Experimental phases yielded high-quality electron density maps that enabled placement of all the amino acids of $Best1_{cryst}$ spanning residues 2–367 and nearly all Fab residues (Extended Data Fig. 3). The asymmetric units contain five ($P2_1$) or ten (C2) BEST1 subunits and a corresponding number of Fab fragments, and the atomic models are refined to crystallographic free residuals of 0.23 and 0.26, respectively, with good stereochemistry (Extended Data Table 1 and Extended Data Fig. 4). Structures of the channels are indistinguishable between the crystal forms (root-mean-square deviation = 0.2 Å). Except where noted, the discussion of the structure pertains to the $P2_1$ crystals, which diffract to higher resolution.

Gating and permeability

We studied the function of purified $BEST1_{cryst}$ by reconstituting it into liposomes and monitoring ion flux using a fluorescence-based assay

[1]Structural Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.
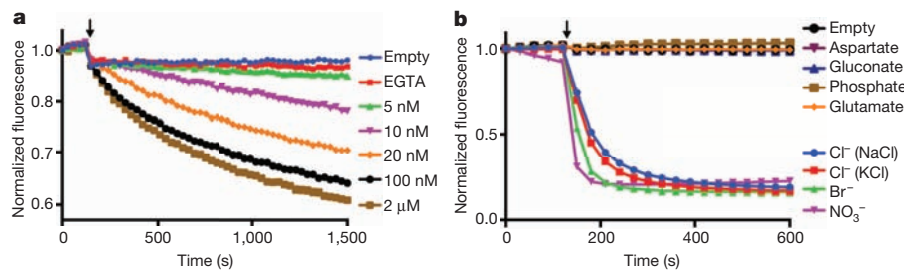
Figure 1 | Gating and ionic permeability of BEST1 in liposomes. a, Purified BEST1$_{cryst}$ recapitulates Ca^{2+}-activated Cl^- flux. Fluorescence traces elicited by various concentrations of free Ca^{2+} are shown. **b**, Anion permeability. Except for KCl, all test ions were sodium salts. The increased rate of

fluorescence decay compared to **a** suggests that channels are predominately oriented with their cytosolic side inside the proteoliposomes. Arrows indicate addition of a proton ionophore. More information on the assay is shown in Extended Data Fig. 5a.

(Fig. 1 and Extended Data Fig. 5). To assay for Ca^{2+}-dependent activation, proteoliposomes that were reconstituted in EGTA and loaded with sulphate, which is essentially impermeant[28], were diluted into solutions containing Cl^- and various concentrations of free Ca^{2+}. We observed fluorescence decreases that depended on the Ca^{2+} concentration indicative of Ca^{2+}-activated permeation of Cl^- into the liposomes (Fig. 1a). Cl^- flux was observed only from liposomes containing BEST1$_{cryst}$ and not from control samples devoid of protein (empty vesicles), and it would result from the fraction of channels that are oriented with their regulatory Ca^{2+} binding site (cytosolic side) facing away from the interior of the liposomes. Ca^{2+}-dependent activation was also observed using NO_3^- as the permeant anion (Extended Data Fig. 5b). To assess the permeability of other anions, BEST1$_{cryst}$ was reconstituted in the presence of ~2 μM free Ca^{2+} to activate channels in both orientations and the proteoliposomes were diluted into solutions containing various test anions. We

observed time-dependent fluorescence decreases indicative of permeability to NO_3^-, Br^- and Cl^- with a permeability sequence of $NO_3^- > Br^- > Cl^-$ (Fig. 1b), which is in agreement with measurements made in cellular contexts[11,28,29,32]. Permeabilities to glutamate, aspartate, gluconate and phosphate were not detected (Supplementary Discussion). Reconstitution of the BEST1$_{cryst}$–Fab complex yielded analogous anion permeation properties to BEST1$_{cryst}$ alone and this indicates that the crystallized complex supports anion flux (Extended Data Fig. 5c). Our results demonstrate that assembled BEST1 oligomers form anion pores that are directly gated by Ca^{2+}.

Architecture

The bestrophin channel is a pentamer of five BEST1 subunits symmetrically arranged around a central axis (Fig. 2 and Supplementary Discussion). It is roughly barrel-shaped with dimensions of ~70 Å across and ~95 Å
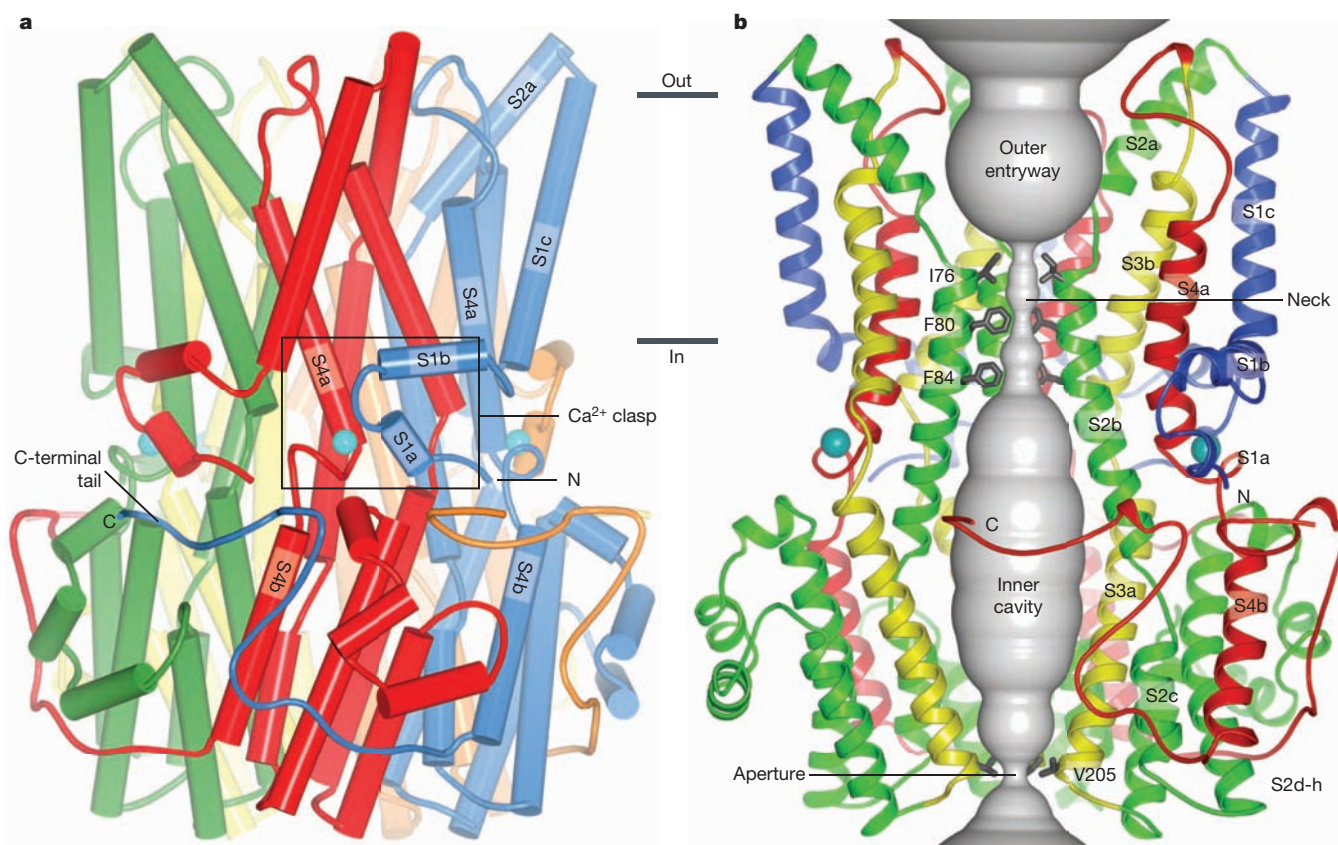
Figure 2 | Architecture and ion pore. a, Overall structure of BEST1$_{cryst}$. The perspective is from within the membrane, with subunits coloured individually, α-helices depicted as cylinders, and approximate boundaries of the membrane indicated. The boxed region highlights a Ca^{2+} clasp with bound Ca^{2+} (teal sphere). **b**, Ion pore. Within a ribbon representation of three

subunits of BEST1 (two in the foreground are removed) is a representation (grey colour) of the minimal radial distance from the centre of the pore to the nearest van der Waals protein contact. Secondary structural elements are coloured according to their four segments (S1, blue; S2, green; S3, yellow; S4 and C-terminal tail, red).

high. A single ion pore is located perpendicular to the plane of the membrane, along the channel's axis of symmetry (Fig. 2b). On the basis of surface hydrophobicity (Extended Data Fig. 6a), the protein extends just beyond the extracellular side of the membrane and protrudes ~55 Å into the cytosol. Five Fab fragments bind with 1:1 stoichiometry to the cytosolic region at a subunit interface and radiate outward (Extended Data Fig. 4a). Each subunit crosses the membrane four times, predominately as α-helices but also as extended conformations (Fig. 2 and Extended Data Fig. 6b). The secondary structure can be divided into four segments accordingly (segments S1–S4). Each segment contributes to the large intracellular region, which appears to be integral to the channel as a whole rather than a domain separate from it (Fig. 2).

Extending from its ordered N terminus (at Thr 2, Methods), the S1 segment runs below the plane of the membrane, forms a lateral helix-turn-helix structural element involving helices S1a and S1b and transitions into the S1c helix that traverses the membrane (Fig. 2). The S1b helix is amphipathic, with hydrophilic amino acids facing the cytosol and hydrophobic amino acids positioned to interact with the lipid membrane. The S1a–S1b helix-turn-helix element is one component of a 'Ca²⁺ clasp' from each subunit that binds intracellular Ca²⁺ (Fig. 2a).

Helices S2a and S2b, which traverse the membrane but are mostly shielded from it, line nearly half of the ion pore (Fig. 2b). The junction between S2a and S2b occurs near the midpoint of the membrane (Tyr 72, Ala 73 and Glu 74) and exposes the N-terminal end of S2b to the pore. Following S2b, six α-helices (S2c–S2h) form a compact structure that comprises the bulk of the intracellular portion of the channel.

The S3 and S4 segments each contain one cytosolic helix (S3a and S4b) and one transmembrane helix (S3b and S4a). S3a and S3b are roughly parallel to S4b and S4a, respectively, and their junctions in secondary structure are similarly located with respect to their positions along the symmetry axis. The amino acids preceding S4a adopt an extended conformation and span approximately one-third of the transmembrane region, leaving the N-terminal end of S4a exposed to the ion pore. The junction between S4a and S4b, which forms a tight turn and contains the highly conserved cluster of acidic amino acids, comprises the other component of the Ca²⁺ clasp. Following S4b, amino acids 326–367 adopt an elongated conformation (the 'C-terminal tail') that wraps around the cytosolic portion of two adjacent subunits (Fig. 2 and Extended Data Fig. 3b). The C-terminal tail is well conserved among bestrophin orthologues (for example, it has the same length and shares 68% amino acid identity with human BEST1) but its sequence is a distinguishing feature of BEST1–4, possibly signifying a modulatory role that imparts functional differences to these paralogues[7] (Supplementary Discussion).

Ion pore

The pore is ~95 Å long and continuous in the sense that there are no lateral openings through which ions might pass. Portions of the S2, S3 and S4 segments line the pore and its diameter varies along its distance (Fig. 2b). An ion moving from the extracellular side towards the intracellular side would encounter a wide funnel-shaped 'outer entryway' (~20 Å across) that narrows to a slender 'neck' near the midpoint of the membrane. The outer entryway is lined by amino acids including those from helix S2a, creating a hydrophilic surface that is exposed to the aqueous extracellular environment. The hydrophobic amino acids Ile 76, Phe 80 and Phe 84 protrude from each of the five S2b helices and line the neck of the pore. Exposure of Phe 80 and other amino acids from the S2 segment to the pore is in agreement with previous studies[33,34]. The region of helix S2b that forms the neck is nearly perpendicular to the membrane plane and angled out slightly such that the neck is perceptively wider at Phe 84, which corresponds approximately to the level of the membrane–cytosol interface (Fig. 2b).

Below Phe 84, the S2b helices bend slightly and the pore opens into a large 'inner cavity' (approximately 45 Å long and 20 Å across at its widest point) that spans the majority of the channel's cytosolic portion before the pore narrows again to its cytosolic 'aperture' (Fig. 2b). Amino acids following the bend in S2b contribute to the surface of the inner cavity,

which is hydrophilic. Tilted S3a helices also line the inner cavity, narrowing it to the aperture at Val 205.

The Ca²⁺ clasp

Electron density consistent with Ca²⁺ was observed within the Ca²⁺ clasp, which consists of the acidic cluster between S4a and S4b (Glu 300, Asp 301, Asp 302, Asp 303 and Asp 304) from one subunit and the S1a–S1b helix-turn-helix element of an adjacent subunit (Fig. 3). The assignment of the electron density to Ca²⁺ is corroborated by the chemistry of coordination and by a corresponding peak in an anomalous difference electron density map (Fig. 3a). To investigate the possibility of additional Ca²⁺ binding site(s), and to determine what effect, if any, the low pH of the P2₁ crystal form has on BEST1, diffraction data were collected from crystals grown in the presence of 5 mM Ca²⁺ at pH 4 and pH 8.5 (P2₁ and C2 forms, respectively), and the atomic models were refined (Extended Data Table 1). No differences in the structure of BEST1 were detected and anomalous difference electron density attributable to Ca²⁺ was observed only in the previously identified Ca²⁺ site.

Together, the five symmetrical Ca²⁺ clasps resemble a belt around the midsection of the channel, below of the membrane–cytosol interface (Fig. 2a). Consistent with a high-affinity interaction, Ca²⁺ is buried by the protein but would become accessible to solvent if S1a–S1b were dislodged. Ca²⁺ coordination has pentagonal bipyramidal geometry, where bidentate coordination by the side chain of Asp 304 along with the backbone carbonyl oxygen atoms of Ala 10 and Gln 293 and an ordered water molecule align along the vertices of an approximately planar pentagon and the side chain of Asp 301 and the backbone carbonyl of Asn 296 take axial positions (Fig. 3b). The coordination is similar to that observed for canonical EF hand domains[35] and for the 'Ca²⁺ bowl' of the BK potassium channel[36], and has an average Ca²⁺–oxygen distance of 2.5 Å. Glu 300, Asp 302 and Asp 303 surround the binding site for Ca²⁺ and although they do not contact the ion directly, they may serve to increase the local concentration of Ca²⁺ (Fig. 3b and Supplementary Discussion). The absence of Ca²⁺ would probably have marked effects on the conformations of the S4a–S4b junction and the S1a–S1b region.

Anion binding

Electron densities at several sites within the ion pore were consistent with bound Cl⁻ ions. To distinguish Cl⁻ from water or other entities, we collected X-ray diffraction data from crystals grown in 150 mM Br⁻, a permeable anion that is crystallographically identifiable from its anomalous X-ray scattering. Anomalous difference electron density
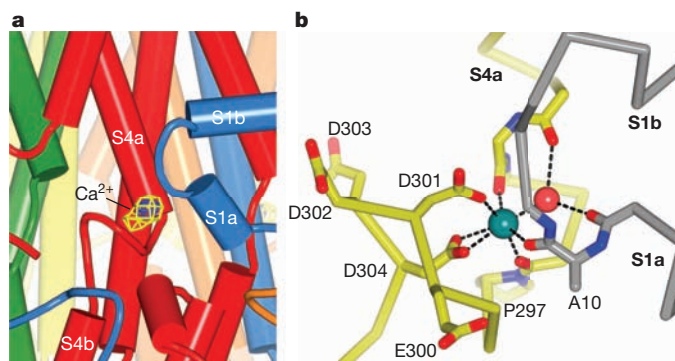
Figure 3 | Ca²⁺ sensing apparatus. a, View of a Ca²⁺ clasp (same orientation as Fig. 2a), showing electron density for Ca²⁺: $F_O − F_C$ density (blue mesh; simulated annealing omit, 40–2.85 Å, 8σ contour) and anomalous difference density (yellow mesh, 40–4.0 Å, 3σ contour). **b,** Coordination in the Ca²⁺ clasp. The acidic cluster and the backbone carbonyls that coordinate (dotted lines) the Ca²⁺ (teal sphere) are depicted as sticks on a Cα representation. Dotted lines also indicate hydrogen bonds between the water molecule (red sphere) and the protein (backbone carbonyls of Val 9 and Glu 292). Carbon atoms of one subunit are grey and those from another are yellow.
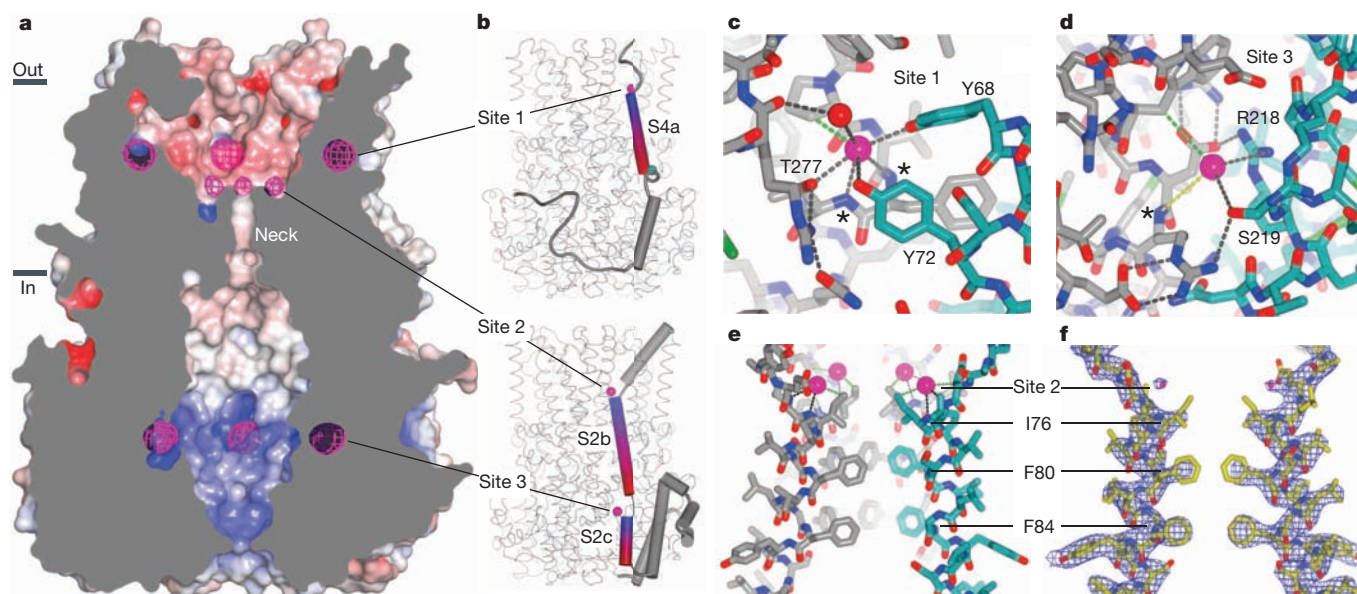
Figure 4 | Anion binding. a, Cut-away view of BEST1, revealing the surface of the pore (coloured by electrostatic potential; red, $-10\,kT\,e^{-1}$; white, neutral; blue, $+10\,kT\,e^{-1}$) and anomalous difference electron density for Br^- ions (magenta mesh; 45–5 Å, non-crystallographic symmetry averaged, 8σ contour). **b**, Anion binding sites (magenta spheres) at the N-terminal ends of α-helices. Representations of the S4 and S2 segments of one subunit (upper and lower panels, respectively) are shown in the context of the entire channel. α-Helices (cylinders) interacting with Cl^-/Br^- are coloured blue-to-red from their N- to their C-terminal ends. A teal sphere (upper panel) denotes Ca^{2+}. **c–e**, Coordination of Cl^- in sites 1, 2 and 3. Interactions (distances <4 Å) with Cl^- (magenta spheres) are shown for polar (grey dashes) and hydrophobic

(green dashes) contacts. Protein is depicted as sticks, with carbon atoms of one subunit coloured teal and those of other subunits grey. Hydrogen bonding networks (in sites 1 and 3) and an ordered water molecule (red sphere in site 1) are shown. In **c** and **d**, asterisks indicate main-chain amide nitrogen atoms at the N-terminal ends of α-helices. A dashed yellow line (**d**) indicates the ~5 Å distance to the N-terminal end of helix S2c. In **e**, Cl^- coordination outside the neck of the pore in site 2 is shown in the context of four S2 segments (foreground segment removed for clarity). **f**, Electron density ($2F_O - F_C$, 40–2.85 Å, 2.0σ contour) for two S2 segments and their corresponding Cl^- ions (magenta spheres) in the same orientation as **e**.

maps indicate the presence of Br^- at three locations within the pore (sites 1–3), with each location exhibiting five-fold symmetry (Fig. 4a and Extended Data Fig. 6c). All of the sites are accessible to the aqueous environment of the pore, with two rings of sites located within the outer entryway (sites 1 and 2) and one ring of sites located within the inner cavity (site 3). Reminiscent of the ClC family of Cl^- channels/transporters[37] and a glutamate-gated Cl^- channel[38], in each of the sites, the Br^-/Cl^- ion is bound adjacent to the N-terminal end of an α-helix where it is stabilized by positive electrostatic potential arising from the oriented peptide dipoles of the helices (Fig. 4b).

The binding in site 1, which is located closest to the extracellular side and at a subunit interface, is stabilized by direct electrostatic interactions with main-chain amide nitrogen atoms at the N-terminal end of helix S4a and by interactions with the side chains of Tyr 68, Tyr 72 from one subunit and Thr 277 of another subunit (Fig. 4c). Electron density consistent with a water molecule, which coordinates the Cl^- ion and is itself stabilized by a hydrogen bond with the protein, delineates an approximate trajectory for the Cl^- into the aqueous environment of the pore (Fig. 4c and Extended Data Fig. 6c).

Site 2 is located at the base of the outer entryway, above the neck, and its position approximately corresponds to the midpoint of the membrane (Fig. 4e). The positive dipole at the end of helix S2b makes the only direct electrostatic interaction with the anion. The absence of other interactions is consistent with the weaker anomalous difference electron density observed at site 2 in comparison to sites 1 and 3, and may be indicative of lower binding affinity.

Site 3 is located within the inner cavity at a subunit interface and is within ~5 Å of the main-chain amide nitrogen of Arg 105 at the N-terminal end of helix S2c from one subunit and is within ~4 Å of the side chains of Arg 218 and Ser 219 from the adjacent subunit (Fig. 4d). Whether these interactions are direct or water-mediated is unclear. Mutations in or around sites 1–3 (for example, Y72D, L75F, I76V, F80L, F84V, R218S) are associated with eye diseases[19].

The observed sites would increase the local concentration of anions on both sides of the neck of the pore and this may contribute to anion selectivity. A similar mechanism has been proposed for an anion-selective Cys-loop receptor[38]. In BEST1, the sites appear well suited for monovalent anions (for example, peptide dipoles in sites 1 and 2 provide the only positive electrostatic potential) and this may contribute to the channel's selectivity for monovalent anions over divalent ones. Except for the positively charged pockets that form the anion binding sites, the electrostatic surface of the outer entryway is predominately negative and it would therefore tend to exclude anions other than the ones that can bind in sites 1 and 2 (Fig. 4a and Extended Data Fig. 6c). The inner cavity is predominately positive and is therefore a favourable environment for anions that can access it.

The permeability sequence of BEST1 for monovalent anions corresponds with their relative hydration energies, which suggests that the ions become at least partially dehydrated at some point during permeation[7,39]. In the neck of the pore, the distances between the central axis and Ile 76, Phe 80 and Phe 84 are approximately 3.8 Å, 3.1 Å and 4.0 Å, respectively (measured to atom CG2 of Ile 76 and to the edge of the phenylalanine rings). The electron densities for Ile 76 and Phe 80 are weaker than for Phe 84 (Fig. 4f), which suggests that there is a degree of 'breathing' of the pore due to side-chain rotamer conformational changes and/or backbone mobility and that the effective diameter of the pore experienced by a permeating anion would be larger than deduced solely from the average positions of these residues. Regardless, an anion passing through the hydrophobic neck would need to be at least partially dehydrated. The relatively low single-channel conductance of bestrophin (~2 pS for *Drosophila* Best1 (ref. 12)) could be due to an energy barrier imposed by the neck. Congruently, although anions are not observed in the neck, they are poised just outside of it.

The aromatic rings of phenylalanine residues have negative electrostatic potential associated with the face of their π system and positive electrostatic potential associated with their edges. Interaction of a cation

with the face of an aromatic ring (the cation-π interaction) has been widely discussed and is important in protein structure and ligand binding (for example, ref. 40). Phe 80 and Phe 84 are positioned such that the edge of each phenylalanine residue interacts with the face of the corresponding phenylalanine from the neighbouring subunit (Extended Data Fig. 7). Such edge–face interactions are commonly observed in proteins. The arrangement is also such that the electrostatically positive edges of the aromatic rings are oriented towards the central axis of the pore (Fig. 4e and Extended Data Fig. 7). This creates positive electrostatic potential along the central axis that could stabilize a permeating anion. The interaction between an anion and the edge of an aromatic ring (the anion–π interaction) is calculated to be energetically favourable and a survey of protein structures indicates that it commonly occurs, for instance where an aspartate interacts with the edge of a phenylalanine[41–43]. On the basis of these studies, the geometries between the central axis of the pore and the aromatic rings of Phe 80 and Phe 84 are favourable for interactions with anions (Extended Data Fig. 7)[41,42]. As such, a permeating anion may interact electrostatically with Phe 80 and Phe 84 within the neck of the pore and this may contribute to anion selectivity.

Retinopathies and the gating apparatus

While mutations associated with eye disease occur in several areas of BEST1, they are particularly prevalent in or around the Ca^{2+} clasp and the neck of the pore (Fig. 5a). This includes mutations of the Ca^{2+} ligands Asp 301 and Asp 304 and the surrounding acidic residues that are known to impair channel function[6,13,26,27], as well as mutations within the S1a–S1b element, consistent with the role of this region in sensing intracellular Ca^{2+}. Mutations within the neck (for example, of Phe 80 and Phe 84) also alter permeation properties of the channel[33,34].

The narrowness of the neck, its high degree of sequence conservation, and its positioning along the pore nearest to the Ca^{2+} binding site suggest that the neck forms a gate. Subtle structural changes near the Ca^{2+} clasp, which we observed between crystals grown using different detergents (Extended Data Fig. 8), are correlated with subtle changes in the diameter of the neck, suggesting that there is conformational coupling between the Ca^{2+} sensor and the gate. We propose that the gate is dilated when Ca^{2+} is bound and seals shut when Ca^{2+} is absent (Fig. 5b). The movements within the gate that switch between conductive and non-conductive states may be limited to side-chain motions or they may be more dramatic. While the Fab does not interact with the Ca^{2+} clasp, its

specificity for the Ca^{2+}-bound form suggests that Ca^{2+}-dependent gating also involves long-range conformational changes that may affect the geometry of the pore in other respects.

Conclusion

The X-ray structure of BEST1 reveals the architecture of a eukaryotic Ca^{2+}-activated Cl^- channel. Crystallized in complex with Ca^{2+} and stabilized by a Fab that preferentially binds the Ca^{2+}-bound form of the channel and supports ion flux, the structure probably represents an open state (or a nearly open state). In several respects, the channel differs in structure and mechanism from other ion channels. Numerous binding sites for Cl^- increase its local concentration and probably contribute to selective permeation. Phenylalanine residues that probably serve as part of the channel's gate may also facilitate anion permeation and contribute to anion selectivity via anion–π interactions. The channel's cytosolic aperture may function as a size-selective filter that permits passage of the small anions permeable to BEST1 while preventing large intracellular anions (for example, proteins and nucleic acids) from accessing the positively charged inner cavity and obstructing the permeation pathway. The gating apparatus, which is often mutated in BEST1-related eye diseases, appears to couple the binding of intracellular Ca^{2+} to dilation of the centrally located ion gate.

1. Hartzell, C., Putzier, I. & Arreola, J. Calcium-activated chloride channels. *Annu. Rev. Physiol.* **67**, 719–758 (2005).
2. Caputo, A. *et al.* TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science* **322**, 590–594 (2008).
3. Schroeder, B. C., Cheng, T., Jan, Y. N. & Jan, L. Y. Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. *Cell* **134**, 1019–1029 (2008).
4. Yang, Y. *et al.* TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature* **455**, 1210–1215 (2008).
5. Sun, H., Tsunenari, T., Yau, K.-W. & Nathans, J. The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proc. Natl Acad. Sci. USA* **99**, 4008–4013 (2002).
6. Tsunenari, T., Nathans, J. & Yau, K.-W. Ca^{2+}-activated Cl^- current from human bestrophin-4 in excised membrane patches. *J. Gen. Physiol.* **127**, 749–754 (2006).
7. Hartzell, H. C., Qu, Z., Yu, K., Xiao, Q. & Chien, L.-T. Molecular physiology of bestrophins: multifunctional membrane proteins linked to best disease and other retinopathies. *Physiol. Rev.* **88**, 639–672 (2008).
8. Kunzelmann, K. *et al.* Role of the Ca^{2+}-activated Cl^- channels bestrophin and anoctamin in epithelial cells. *Biol. Chem.* **392**, 125–134 (2011).
9. Gomez, N. M., Tamm, E. R. & Straubeta, O. Role of bestrophin-1 in store-operated calcium entry in retinal pigment epithelium. *Pflugers Archiv.* **465**, 481–495 (2013).
10. Tsunenari, T. *et al.* Structure-function analysis of the bestrophin family of anion channels. *J. Biol. Chem.* **278**, 41114–41125 (2003).
11. Qu, Z., Fischmeister, R. & Hartzell, C. Mouse bestrophin-2 is a bona fide Cl^- channel: identification of a residue important in anion binding and conduction. *J. Gen. Physiol.* **123**, 327–340 (2004).
12. Chien, L.-T., Zhang, Z.-R. & Hartzell, H. C. Single Cl^- channels activated by Ca^{2+} in *Drosophila* S2 cells are mediated by bestrophins. *J. Gen. Physiol.* **128**, 247–259 (2006).
13. Xiao, Q., Prussia, A., Yu, K., Cui, Y.-y. & Hartzell, H. C. Regulation of bestrophin Cl channels by calcium: role of the C terminus. *J. Gen. Physiol.* **132**, 681–692 (2008).
14. Marquardt, A. *et al.* Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). *Hum. Mol. Genet.* **7**, 1517–1525 (1998).
15. Petrukhin, K. *et al.* Identification of the gene responsible for Best macular dystrophy. *Nature Genet.* **19**, 241–247 (1998).
16. Davidson, A. E. *et al.* Missense mutations in a retinal pigment epithelium protein, Bestrophin-1, cause retinitis pigmentosa. *Am. J. Hum. Genet.* **85**, 581–592 (2009).
17. Boon, C. J. *et al.* The spectrum of ocular phenotypes caused by mutations in the BEST1 gene. *Prog. Retin. Eye Res.* **28**, 187–205 (2009).
18. Xiao, Q., Hartzell, H. C. & Yu, K. Bestrophins and retinopathies. *Pflügers Archiv.* **460**, 559–569 (2010).
19. Kinnick, T. R. *et al.* Autosomal recessive vitelliform macular dystrophy in a large cohort of vitelliform macular dystrophy patients. *Retina* **31**, 581–595 (2011).
20. Wittström, E., Ponjavic, V., Bondeson, M. L. & Andreasson, S. Anterior segment abnormalities and angle-closure glaucoma in a family with a mutation in the BEST1 gene and Best vitelliform macular dystrophy. *Ophthalmic Genet.* **32**, 217–227 (2011).

Figure 5 | Retinopathies and the gating apparatus. a, Locations of missense mutations associated with retinal diseases[19] mapped on the structure (red spheres indicate Cα positions). Teal spheres represent Ca^{2+}. **b**, Hypothesized mechanisms of gating and selectivity. Intracellular Ca^{2+} binding is coupled to dilation of the gate (neck). Within the context of the otherwise negatively charged outer entryway, binding sites for monovalent anions (magenta) increase their local concentration. Phenylalanine residues within the gate may contribute to selective anion permeation via anion–π interactions ($δ^+$). Additional binding sites for anions are located in the predominately positive inner cavity.

21. Yu, K., Cui, Y. & Hartzell, H. C. The bestrophin mutation A243V, linked to adult-onset vitelliform macular dystrophy, impairs its chloride channel function. *Invest. Ophthalmol. Vis. Sci.* **47,** 4956–4961 (2006).

22. Yu, K., Qu, Z., Cui, Y. & Hartzell, H. C. Chloride channel activity of bestrophin mutants associated with mild or late-onset macular degeneration. *Invest. Ophthal. Vis. Sci.* **48,** 4694–4705 (2007).

23. Marchant, D. *et al.* New VMD2 gene mutations identified in patients affected by Best vitelliform macular dystrophy. *J. Med. Genet.* **44,** e70 (2007).

24. Milenkovic, V. M., Röhrl, E., Weber, B. H. F. & Strauss, O. Disease-associated missense mutations in bestrophin-1 affect cellular trafficking and anion conductance. *J. Cell Sci.* **124,** 2988–2996 (2011).

25. Bharill, S., Fu, Z., Palty, R. & Isacoff, E. Y. Stoichiometry and specific assembly of Best ion channels. *Proc. Natl Acad. Sci. USA* **111,** 6491–6496 (2014).

26. Qu, Z. Q., Yu, K., Cui, Y.-y. & Ying, C. & Hartzell, C. Activation of bestrophin Cl⁻ channels is regulated by C-terminal domains. *J. Biol. Chem.* **282,** 17460–17467 (2007).

27. Kranjc, A. *et al.* Regulation of bestrophins by Ca²⁺: a theoretical and experimental study. *PLoS ONE* **4,** e4672 (2009).

28. Qu, Z. & Hartzell, H. C. Bestrophin Cl⁻ channels are highly permeable to HCO₃⁻. *Am. J. Physiol. Cell Physiol.* **294,** C1371–C1377 (2008).

29. Stotz, S. C. & Clapham, D. E. Anion-sensitive fluorophore identifies the *Drosophila* swell-activated chloride channel in a genome-wide RNA interference screen. *PLoS ONE* **7,** e46865 (2012).

30. Lee, S. *et al.* Channel-mediated tonic GABA release from glia. *Science* **330,** 790–796 (2010).

31. Woo, D. H. *et al.* TREK-1 and Best1 channels mediate fast and slow glutamate release in astrocytes upon GPCR activation. *Cell* **151,** 25–40 (2012).

32. O'Driscoll, K. E., Leblanc, N., Hatton, W. J. & Britton, F. C. Functional properties of murine bestrophin 1 channel. *Biochem. Biophys. Res. Commun.* **384,** 476–481 (2009).

33. Qu, Z. & Hartzell, C. Determinants of anion permeation in the second transmembrane domain of the mouse bestrophin-2 chloride channel. *J. Gen. Physiol.* **124,** 371–382 (2004).

34. Qu, Z., Chien, L.-T., Cui, Y. & Hartzell, H. C. The anion-selective pore of the bestrophins, a family of chloride channels associated with retinal degeneration. *J. Neurosci.* **26,** 5411–5419 (2006).

35. Gifford, J. L., Walsh, M. P. & Vogel, H. J. Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs. *Biochem. J.* **405,** 199–221 (2007).

36. Yuan, P., Leonetti, M. D., Hsiung, Y. & MacKinnon, R. Open structure of the Ca²⁺ gating ring in the high-conductance Ca²⁺-activated K⁺ channel. *Nature* **481,** 94–97 (2012).

37. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415,** 287–294 (2002).

38. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474,** 54–60 (2011).

39. Hille, B. *Ionic Channels of Excitable Membranes* 2nd edn (Sinauer Associates, 1992).

40. Dougherty, D. A. Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271,** 163–168 (1996).

41. Jackson, M. R. *et al.* A preference for edgewise interactions between aromatic rings and carboxylate anions: the biological relevance of anion-quadrupole interactions. *J. Phys. Chem. B* **111,** 8242–8249 (2007).

42. Philip, V. *et al.* A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-π pairs. *Biochemistry* **50,** 2939–2950 (2011).

43. Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldmann, R. J. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proc. Natl Acad. Sci. USA* **79,** 4843–4847 (1982).

METHODS

Cloning, expression and purification of BEST1$_{cryst}$. Chicken (*Gallus gallus*) bestrophin 1 was cloned from cDNA (BioChain) and identified as a promising candidate for protein purification and crystallization from among 30 eukaryotic orthologues of human bestrophin 1 that we evaluated using the fluorescence-detection size exclusion chromatography (FSEC) pre-crystallization screening technique[44]. Guided by sequence conservation, limited proteolysis of purified protein, and predicted secondary structure, a construct spanning amino acids 1–405 of chicken BEST1 was used for crystallization (BEST1$_{cryst}$). cDNA encoding BEST1$_{cryst}$ was cloned into pPICZ (Invitrogen) and consists of amino acids 1–405 followed by an affinity tag (Glu-Gly-Glu-Glu-Phe) that is recognized by an anti-tubulin antibody (designated YL½)[45]. Transformation into *Pichia pastoris*, protein expression and lysis was performed as previously described[46].

Lysed cells were resuspended (using ~10 ml of buffer for each gram of cells) in a purification buffer consisting of 50 mM Tris-HCl, pH 7.5, 75 mM NaCl, 75 mM KCl, 0.1 mg ml^{-1} DNase I (Sigma-Aldrich), a 1:600 dilution of Protease Inhibitor Cocktail Set III, EDTA-free (CalBiochem), and 0.5 mM 4-(2-aminoethyl) benzenesulphonyl fluoride hydrochloride (Gold Biotechnology). 0.14 g of *n*-dodecyl-β-D-maltopyranoside (DDM; Anatrace) was added per 1 g of cells, the pH was adjusted to pH 7.5 using 1 M NaOH, and the sample was agitated for 45 min at room temperature. Following extraction, the sample was clarified by centrifugation at 43,000*g* at 12 °C for 40 min and filtered using a 0.45 μm polyethersulphone membrane. Affinity purification was achieved using YL½ antibody (IgG, expressed by hybridoma cells and purified by ion exchange chromatography) that was coupled to CNBr-activated sepharose beads according to the manufacturer's protocol (GE Healthcare). 1.0–2.0 ml of resin was added to the sample for each 1 g of *P. pastoris* cell lysate and the mixture was rotated at room temperature for 1 h. The mixture was then applied to a column support and was washed with ~5 column volumes of a buffer containing 20 mM Tris-HCl, pH 7.5, 75 mM NaCl, 75 mM KCl and 3 mM DDM. Elution was carried out using 4 column volumes of elution buffer: 100 mM Tris-HCl, pH 7.5, 75 mM NaCl, 75 mM KCl, 3 mM DDM and 5 mM Asp-Phe peptide (Sigma-Aldrich). The elution fraction was concentrated to ~2 mg ml^{-1} using a 100,000 Da concentrator (Amicon Ultra; EMD Millipore) before combining with the Fab. Mass spectrometry and Edman degradation of purified BEST1$_{cryst}$ indicate that the initial methionine has been removed and that the amino terminus is at Thr 2.

Fab production and co-crystallization. A monoclonal antibody (designated 10D10) of isotype IgG1 was raised in mice by the Monoclonal Antibody Core Facility of the Memorial Sloan Kettering Cancer Center and selected for co-crystallization with BEST1$_{cryst}$. The antigen used for immunization was BEST1$_{cryst}$ that had been purified in DDM and digested using the serine protease GluC (Worthington), which removes approximately 20 amino acids from the C terminus of the protein. The antibody selection process included ELISA, western blot, and FSEC analysis to identify antibodies that bound to native BEST1$_{cryst}$ and not SDS-denatured protein. The cDNA sequence of the antibody was determined from hybridoma cells by SYD Labs. The antibody was expressed using mouse hybridoma cells, purified by ion exchange chromatography and cleaved using papain (Worthington) to generate the Fab fragment. The Fab fragment was purified using ion exchange chromatography (Mono S, GE Healthcare), dialysed into buffer containing 20 mM Tris-HCl, pH 7.5, 75 mM NaCl, 75 mM KCl, and further purified using size exclusion chromatography (SEC) (Superdex-200 10/300 GL, GE Healthcare) in the same buffer immediately before combination with BEST1$_{cryst}$. The purification buffers contained approximately 1 μM Ca^{2+}, which was present due to impurities and was determined using the Fura-2 calcium indicator (Invitrogen). The protein preparations of BEST1$_{cryst}$ and Fab (~2 mg ml^{-1}) were combined in a molar ratio of 1:1.2 (BEST1$_{cryst}$:Fab) such that the concentration of DDM was ~1.5 mM, concentrated using a 10-kDa molecular weight cutoff concentrator (Vivaspin 15R, Sartorius) to ~15 mg ml^{-1} and purified using SEC. The SEC buffer contained 10 mM Tris, pH 7.5, 75 mM NaCl, 75 mM KCl, and one of the following three detergents: (1) 3 mM 6-cyclohexyl-1-hexyl-β-D-maltopyranoside (cymal-6; Anatrace); (2) 0.5 mM 2,2-bis (3'-cyclohexylbutyl) propane-1,3-bis-β-D-maltopyranoside (cymal-6-NG; Anatrace); or (3) 5 mM *n*-decyl-β-D-maltopyranoside (DM; Anatrace). For crystallization with Br$^-$, 150 mM NaBr was used in place of NaCl and KCl. The elution fraction containing the BEST1$_{cryst}$–Fab complex was concentrated to ~14 mg ml^{-1} using a 100 kDa concentrator (Amicon Ultra; EMD Millipore). 50 mM GABA was then added as a crystallization additive and the sample (at ~12 mg ml^{-1}) was used for crystallization trials. GABA improved the reliability of obtaining well-formed crystals but was not required for crystallization. For crystallization with additional Ca^{2+}, 5 mM CaCl$_2$ was added to the sample before crystallization. Crystals formed in the absence of the Fab but were pathological (poor diffraction, severe anisotropy and crystal twinning).

BEST1$_{cryst}$–Fab crystals belonging to the *P2$_1$* space group were obtained using vapour diffusion from protein that was purified in cymal-6 or cymal-6-NG (1:1 ratio of protein:crystallization solution) using a crystallization solution of 0–60 mM

NaCl, 50 mM sodium acetate, pH 4.0, 5% (w/v) PEG 4000, and 20% (v/v) glycerol at 20 °C. These crystals were harvested after 5–10 days and flash-cooled in liquid nitrogen. Crystals belonging to the *C2* space group were grown by vapour diffusion (1:1 ratio of protein to crystallization solution) using a crystallization solution of 120 mM NaCl, 50 mM Tris, pH 8.5, 8.5% PEG 4000, and 20% glycerol at 25 °C. The crystals were harvested using nylon loops and transferred in a series of five steps to increase the PEG 4000 to 25% before flash-cooling in liquid nitrogen. Diffraction data were collected from crystals cooled at 100 K under a stream nitrogen gas using Pilatus 6M detectors (Dectris) at Brookhaven National Synchrotron Light Source (beamline X25) or the Advanced Photon Source (beamline 24-ID-C).

Structure determination. Initial phases (50–6 Å) were determined using a tantalum bromide-derivatized crystal belonging to the *P2$_1$* space group via the SAD method in SHARP[47] (Extended Data Table 1, anomalous phasing power = 1.3 from 50–6 Å and 0.78 in the 6.1–6.0 Å shell). To prepare the tantalum bromide derivative (*P2$_1$* form), solid (Ta$_6$Br$_{12}$)Br$_2$ (Jena Bioscience) was added to crystallization drops containing suitable crystals, and these were incubated at 20 °C for 2 days followed by another addition of solid (Ta$_6$Br$_{12}$)Br$_2$ and further incubation for 3 days. The 'native' *C2* crystal was also incubated with a smaller amount of solid (Ta$_6$Br$_{12}$)Br$_2$ for 24 h, but no evidence of tantalum bromide could be detected in electron density maps. Diffraction data were collected using an oscillation angle of ~0.3° and high redundancy was permitted by collecting data from multiple locations throughout the crystals. Diffraction data were processed with HKL3000[48] and resolution limits were assessed using the CC$_{1/2}$ statistic[49].

Phases were extended and improved using solvent flattening, histogram matching, and five-fold non-crystallographic symmetry (NCS) averaging with the program DM[50] (yielding a figure of merit = 0.82 for the resolution range 50–4.4 Å and figure of merit = 0.78 for 4.5–4.4 Å shell). An atomic model was built using the coot and O software programs[51] and improved through iterative cycles of refinement (using CNS, Refmac, and PHENIX[52–54]), making use of bulk solvent, NCS and TLS refinement strategies. Electron density is continuous for BEST1$_{cryst}$ residues 2–367 and also clear for the Fabs. Initial phases for diffraction data collected from crystals belonging to the *C2* space group were determined by molecular replacement (PHENIX[53]). The atomic model required slight rigid body adjustments to the constant immunoglobulin domains of the Fabs and it was refined in PHENIX, making use of the tenfold non-crystallographic symmetry. Comprehensive model validation was performed with MolProbity[55] (within PHENIX). Data collection and refinement statistics are shown as Extended Data Table 1. Molecular graphics figures were prepared using the programs PyMOL (http://www.pymol.org/) with the APBS plugin[56] and using the program HOLE[57].

Anion flux assay. For reconstitution into liposomes, BEST1$_{cryst}$ was purified as described above except that SEC was performed in the absence of the Fab and the SEC buffer consisted of 150 mM NaCl, 20 mM Tris-HCl, pH 8.5, and 3 mM *n*-decyl-β-D-maltopyranoside (DM). The reconstitution procedure was based on methods described previously[58]. A 3:1 (wt:wt) mixture of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; Avanti) and POPG (1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol; Avanti) lipids was prepared at 20 mg ml^{-1} in one of the two reconstitution buffers indicated below and the lipids were solubilized with 8% *n*-octyl-β-D-maltopyranoside (Anatrace). The protein was then mixed with an equal volume of the solubilized lipids to give a final protein concentration of 0.1 mg ml^{-1} and a lipid concentration of 10 mg ml^{-1}. Detergent was removed by dialysis (8,000 Da molecular mass cutoff) at 4 °C against a total of 10 l of reconstitution buffer with daily buffer exchanges over a course of 5 days. For the ion permeability experiments (Fig. 1b and Extended Data Fig. 5c), 10 μM CaCl$_2$ was added to the protein following SEC and the reconstitution buffer consisted of: 100 mM sodium sulphate, 0.2 mM EGTA, 0.19 mM CaCl$_2$ and 10 mM HEPES, where the pH was adjusted to 7.0 using NaOH. The free Ca^{2+} concentration of this buffer was ~2 μM as determined using the Fura-2 calcium indicator (Invitrogen). For Ca^{2+} gating experiments (Fig. 1a and Extended Data Fig. 5b), purified protein was used without the addition of CaCl$_2$ and the reconstitution buffer was: 100 mM sodium sulphate, 1 mM EGTA, 10 mM HEPES, and the pH was adjusted to 8.1 with NaOH. The higher pH of this buffer was necessary to sufficiently chelate Ca^{2+} using EGTA to close the channel. 'Empty' (lipid only) vesicles were prepared in parallel in the same manner in the absence of protein. Following dialysis, the liposomes were sonicated for approximately 20 s in a water bath, divided into aliquots, and flash-frozen in liquid nitrogen for storage at −80 °C.

Reconstitution of the BEST1$_{cryst}$–Fab complex (Extended Data Fig. 5c) was done in parallel using the same preparation of BEST1$_{cryst}$ and using the same reconstitution buffer (100 mM sodium sulphate, 0.2 mM EGTA, 0.19 mM CaCl$_2$ and 10 mM HEPES-NaOH, pH 7.0). For this, BEST1$_{cryst}$ and Fab were combined in SEC buffer supplemented with 10 μM CaCl$_2$ to yield an excess of Fab (BEST1$_{cryst}$: Fab molar ratio of approximately 1:1.7). The sample was then mixed with an equal volume of solubilized lipids to give a final BEST1$_{cryst}$ concentration of 0.1 mg ml^{-1}, a Fab concentration of 0.18 mg ml^{-1} and a lipid concentration of 10 mg ml^{-1}.

Proteoliposomes were then produced in the same manner as the sample without Fab. Prior to combining the sample with solubilized lipids, a small amount of the sample was analysed by SEC (in 150 mM NaCl, 20 mM Tris-HCl, pH 8.5, and 3 mM DM) in comparison to the analogous sample of $BEST1_{cryst}$ alone and to the Fab. A shift in the elution volume (13.1 ml for $BEST1_{cryst}$ and 12.3 ml for the $BEST1_{cryst}$–Fab complex using a Superdex-200 10/300 GL column) and quantification of the amount of free Fab confirmed formation of the $BEST1_{cryst}$–Fab complex before reconstitution into liposomes. To evaluate whether the $BEST1_{cryst}$–Fab complex was intact in the proteoliposomes, the amount of unbound Fab was quantified following dialysis by SEC (using reconstitution buffer as the running buffer, without detergent) and it was determined to be the same as the amount of excess Fab (within error) before reconstitution. If the Fab had dissociated from $BEST1_{cryst}$ as a result of the reconstitution of $BEST1_{cryst}$ into liposomes, then the amount of excess Fab would be more than twice its initial value and therefore, to a first approximation, the complex was fully intact in the proteoliposomes.

The flux assay was based on previously published methods[58–60]. Vesicles were thawed in a 37 °C water bath, sonicated (for ~30 s, in 10-s intervals), and diluted by 100-fold into a flux assay buffer. For ion permeability experiments (Fig. 1b and Extended Data Fig. 5c), the flux assay buffer consisted of 10 mM HEPES-NaOH, pH 7.0, 0.2 mM EGTA, 0.19 mM $CaCl_2$, 0.5 mg ml^{-1} bovine serum albumin (BSA), 2 μM 9-amino-6-chloro-2-methoxyacridine (ACMA, Sigma-Aldrich, from a 2 mM stock solution in DMSO), and a test salt. The free Ca^{2+} concentration was ~2 μM (determined using Fura-2). The test salts used were: 125 mM NaCl, 125 mM KCl, 125 mM NaBr, 125 mM $NaNO_3$, 125 mM sodium L-aspartate, 125 mM sodium L-glutamate, 110 mM sodium D-gluconate, or a mixture of Na_2HPO_4 and NaH_2PO_4 containing 110 mM phosphate to yield a pH of 7.0. Test salt concentrations were chosen to yield flux assay buffers with approximately the same osmolality as the reconstitution buffer (~255 mOsm, Vapro 5600 osmometer; Wescor Biomedical Systems). Data were collected on a SpectraMax M5 fluorometer (Molecular Devices) using the Softmax Pro 5 software package. Fluorescence intensity measurements were collected every 30 s with excitation and emission wavelengths of 410 nm and 490 nm, respectively. 1 μM of the proton ionophore carbonyl cyanide m-chlorophenyl hydrazone (CCCP, Sigma-Aldrich, from a 1 mM stock solution in DMSO) was added after 120 s and the sample was gently mixed with a pipette in advance of the reading at the 150 s time point. Fluorescence readings were normalized by dividing by the initial reading and were comparable before normalization. Experiments using $BEST1_{cryst}$ (Fig. 1b) and the $BEST1_{cryst}$–Fab complex (Extended Data Fig. 5c) were recorded in parallel on the same day and using the same solutions. The trace for the empty vesicle control (Fig. 1b and Extended Data Fig. 5c) shows results using $NaNO_3$ and is representative of results obtained using other salts.

For Ca^{2+} gating experiments, the flux assay buffer consisted of 125 mM NaCl (Fig. 1a) or 125 mM $NaNO_3$ (Extended Data Fig. 5b) and 10 mM HEPES-NaOH, pH 8.1, 0.5 mg ml^{-1} BSA, 2 μM ACMA, and mixtures of 1 mM EGTA and 1 mM Ca-EGTA to yield a range of free $[Ca^{2+}]$. A Ca-EGTA stock solution was made by mixing 95 mM $CaCO_3$ and 100 mM EGTA at pH 8.1 (adjusted with NaOH) and titrating the final $[Ca^{2+}]$ using $CaSO_4$ to make it equal to [EGTA] by the pH-metric method[61]. The concentrations of free $[Ca^{2+}]$ were calculated using Chelator[62] as implemented at http://maxchelator.stanford.edu/CaEGTA-TS.htm. Experiments in Fig. 1a and Extended Data Fig. 5b were recorded on the same day using the same

batch of proteoliposomes. Traces shown for empty vesicles (2 μM free Ca^{2+}) are representative of other Ca^{2+} concentrations.

Fab binding assay. To assess binding of the Fab to $BEST1_{cryst}$ (Extended Data Fig. 2), 8 nM Fab was incubated (>30 min at 4 °C) with various concentrations of $BEST1_{cryst}$ ranging from 10 nM to 600 nM in buffer (75 mM NaCl, 75 mM KCl, 1 mM DDM, 20 mM Tris-HCl at pH 8.5) containing either 5 mM EGTA or 10 μM $CaCl_2$. 400 μl of each mixture was loaded onto an SEC column (Superdex-200 10/300 GL), which was equilibrated in the same buffer, and the fraction of unbound Fab was quantified from the area under the elution peak corresponding to free Fab (at 17.3 ml and detected using tryptophan fluorescence on a Shimazdu RF-20AXS fluorescence detector), which is well separated from the peaks for $BEST1_{cryst}$ and the $BEST1_{cryst}$–Fab complex (13.1 ml and 12.3 ml, respectively), in comparison to a Fab control.

44. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14,** 673–681 (2006).
45. Kilmartin, J. V., Wright, B. & Milstein, C. Rat monoclonal antitubulin antibodies derived by using a new nonsecreting rat cell line. *J. Cell Biol.* **93,** 576–582 (1982).
46. Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K$^+$ channel. *Science* **309,** 897–903 (2005).
47. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
48. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62,** 859–866 (2006).
49. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
50. Cowtan, K. D. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **31,** 34–38 (1994).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
52. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
53. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
54. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
55. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
56. Konecny, R., Baker, N. A. & McCammon, J. A. iAPBS: a programming interface to Adaptive Poisson-Boltzmann Solver (APBS). *Comp. Sci. Discov.* **5,** 015005 (2012).
57. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14,** 354–360,–376 (1996).
58. Miller, A. N. & Long, S. B. Crystal structure of the human two-pore domain potassium channel K2P1. *Science* **335,** 432–436 (2012).
59. Hou, X., Pedi, L., Diver, M. M. & Long, S. B. Crystal structure of the calcium release-activated calcium channel Orai. *Science* **338,** 1308–1313 (2012).
60. Lee, S.-Y., Letts, J. A. & MacKinnon, R. Functional reconstitution of purified human Hv1 H$^+$ channels. *J. Mol. Biol.* **387,** 1055–1060 (2009).
61. Tsien, R. & Pozzan, T. Measurement of cytosolic free Ca^{2+} with quin2. *Methods Enzymol.* **172,** 230–262 (1989).
62. Schoenmakers, T. J., Visser, G. J., Flik, G. & Theuvenet, A. P. CHELATOR: an improved method for computing metal ion concentrations in physiological solutions. *Biotechniques* **12,** 870–874,–876–879 (1992).
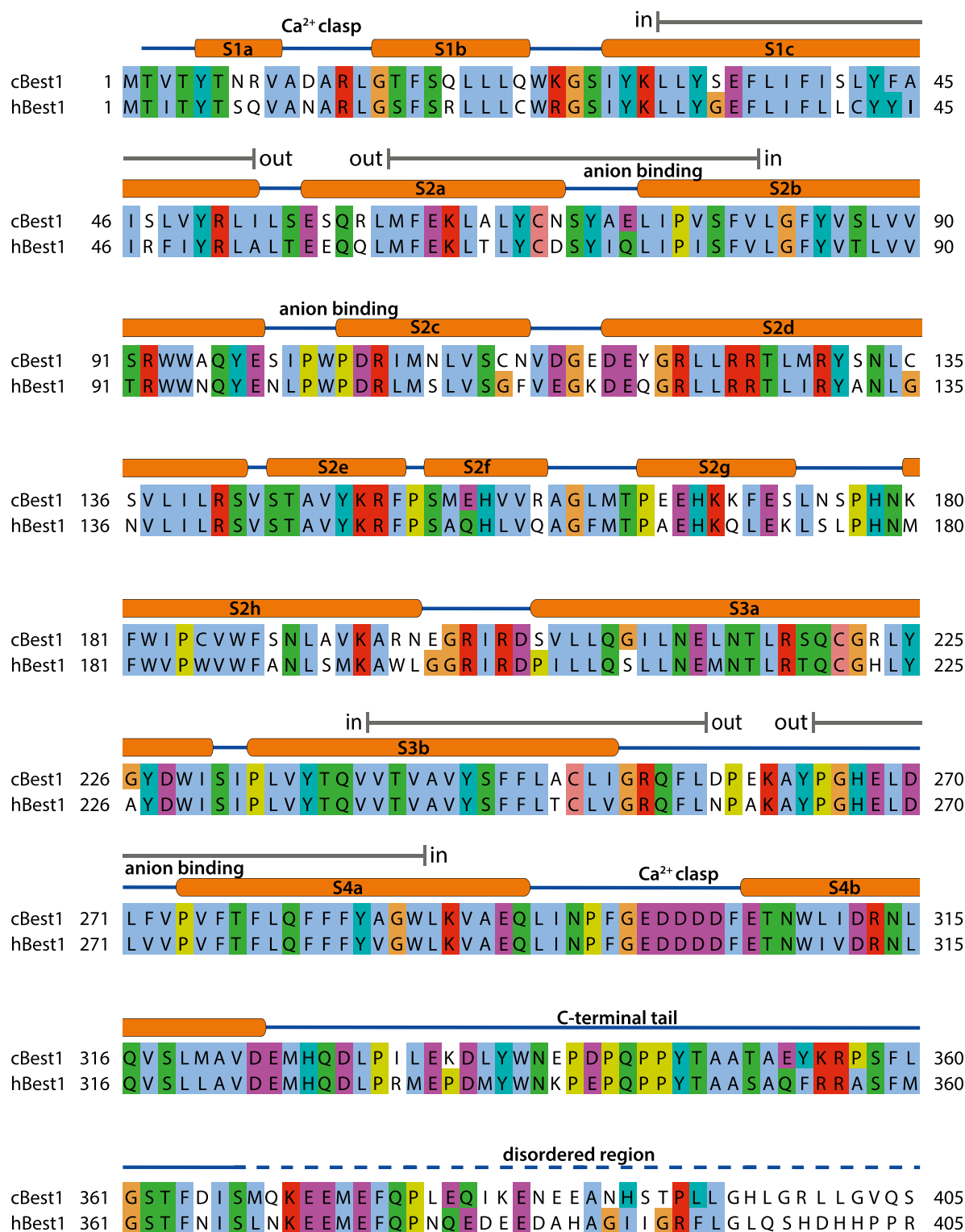
Extended Data Figure 1 | Sequence alignment and secondary structure. The amino acid sequences of the crystallized chicken (*Gallus gallus*) BEST1 construct (amino acids 2–405) and human BEST1 are aligned and coloured according to the ClustalW convention. The secondary structure is indicated with cylinders representing α-helices, solid lines representing structured loop regions, and dashed lines representing disordered regions. Grey bars (labelled 'in' and 'out') indicate approximate boundaries of transmembrane regions.
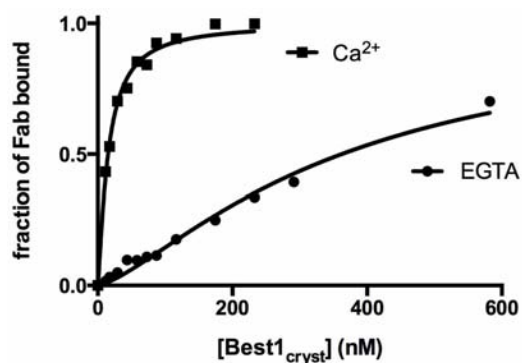
Extended Data Figure 2 | Fab binding to BEST1$_{cryst}$ in the presence and absence of Ca^{2+}. The binding of the Fab to BEST1$_{cryst}$ was assayed by determining the amount of free Fab as a function of the concentration of BEST1$_{cryst}$ in the presence of either 10 μM Ca^{2+} or 5 mM EGTA (zero Ca^{2+}) (Methods). The fraction of Fab bound is plotted with respect to the concentration of BEST1$_{cryst}$. The curves correspond to fits of: fraction of Fab bound = [BEST1]^h/(K_d^h + [BEST1]^h), where K_d is the equilibrium dissociation constant, h is the Hill coefficient, and [BEST1] is the BEST1$_{cryst}$ concentration. Derived parameters are: K_d = 15 nM in the presence of Ca^{2+} (h = 1.3) and K_d = 350 nM in the absence of Ca^{2+} (h = 1.3).
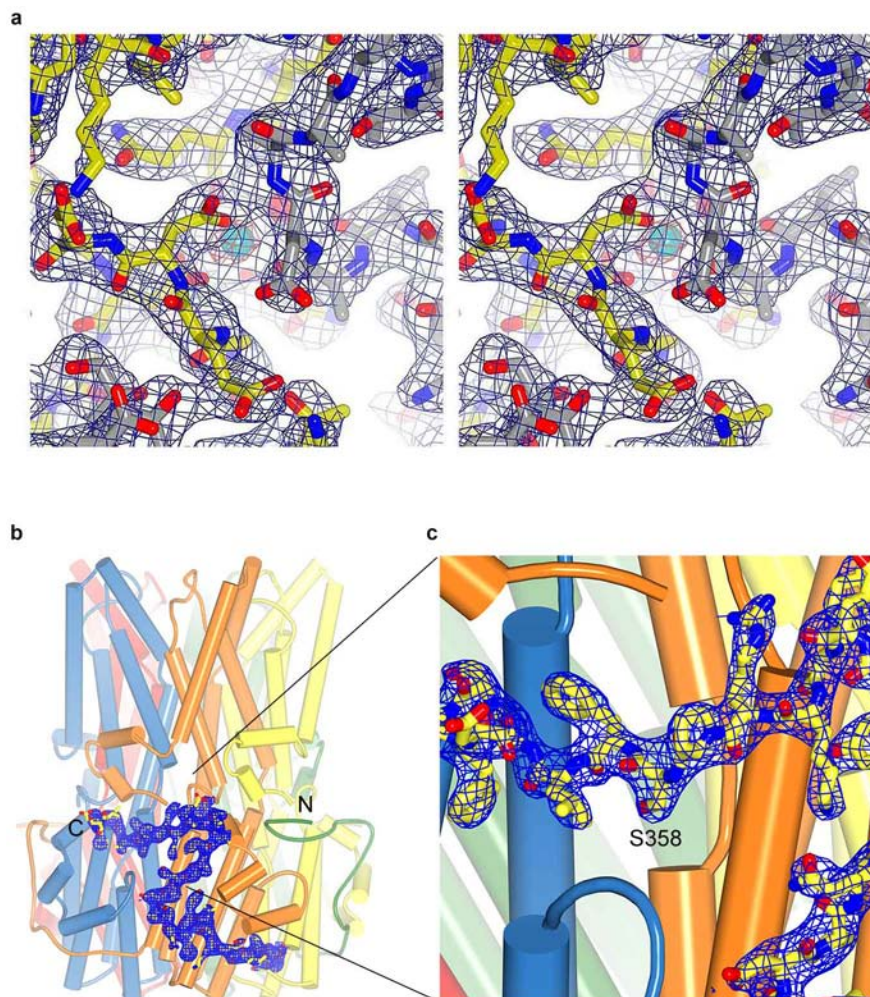
Extended Data Figure 3 | Electron density and the C-terminal tail.
a, $2F_o - F_c$ electron density is shown, in stereo, for an area surrounding one of the five identical Ca^{2+} binding sites. The density was calculated from 40 to 2.85 Å resolution and contoured at 1.5σ (blue mesh) and 7σ (orange mesh) in the context of the final atomic model, which is shown as sticks and spheres (cyan sphere, calcium; red sphere, water). **b**, Electron density for the C-terminal tail. $2F_o - F_c$ electron density (blue mesh, calculated from 40 to 2.85 Å, and contoured at 1.5σ) is shown for the C-terminal tail of the yellow coloured subunit. **c**, Expanded view highlighting the electron density near Ser 358. Consistent with the electron density, mass spectrometry analysis of tryptic peptides of purified BEST1$_{cryst}$ detected only peptides containing Ser 358 that were not phosphorylated (Supplementary Discussion).
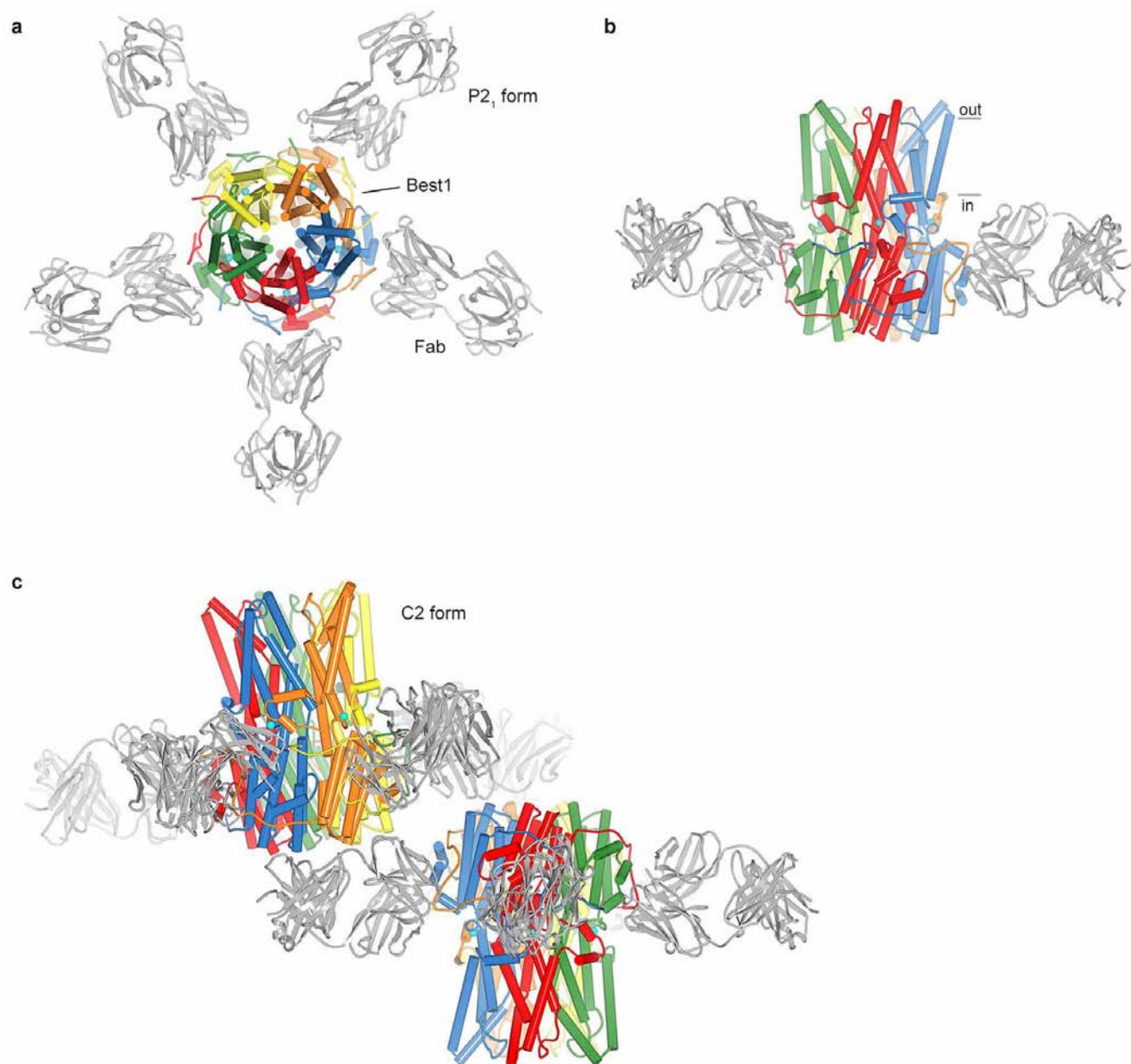
Extended Data Figure 4 | Overall structures of the BEST1$_{cryst}$–Fab complex. a, Structure of the BEST1$_{cryst}$–Fab complex in the $P2_1$ crystal form, viewed from the extracellular side. Fab molecules are grey and BEST1 subunits are coloured individually with α-helices depicted as cylinders. **b,** Orthogonal view showing approximate boundaries of the membrane. For clarity, two Fabs are drawn. **c,** C2 crystal form. Overall structures of the two BEST1$_{cryst}$–Fab complexes in the asymmetric unit of the C2 crystal form are depicted in cartoon representation. BEST1 subunits are coloured individually and Fabs are grey.

Extended Data Figure 5 | Ca^{2+}-dependent activation of $Best1_{cryst}$ and permeability of the $BEST1_{cryst}$–Fab complex. a, Schematic of the fluorescence-based flux assay. Vesicles diluted into various test salts establish ion gradients. Anion influx through BEST1 produces a negative electric potential within the liposomes that drives the uptake of protons through an ionophore (CCCP) and quenches the fluorescence of a pH indicator (ACMA). **b**, Ca^{2+}-dependent activation of $BEST1_{cryst}$ using NO_3^- as the permeant anion. The experimental setup was identical to that for Fig. 1a, except that NO_3^- was used as the permeant ion. Data presented here and in Fig. 1a were collected on the same day using the same batch of proteoliposomes and indicate the higher permeability of NO_3^- relative to Cl^-. Free concentrations of Ca^{2+} are indicated. **c**, Ionic permeability of the $BEST1_{cryst}$–Fab complex. The experiment setup is identical to that shown in Fig. 1b, except that it was performed using proteoliposomes reconstituted with the $BEST1_{cryst}$–Fab complex. The Fab remained bound to the channel following reconstitution and excess Fab was maintained throughout (Methods). The slight differences in the shape of the curves for the $BEST1_{cryst}$ and $BEST1_{cryst}$–Fab samples (for example, the lower rate of fluorescence decrease for Cl^- compared with Fig. 1b) are in accord with variability observed among different liposome preparations.
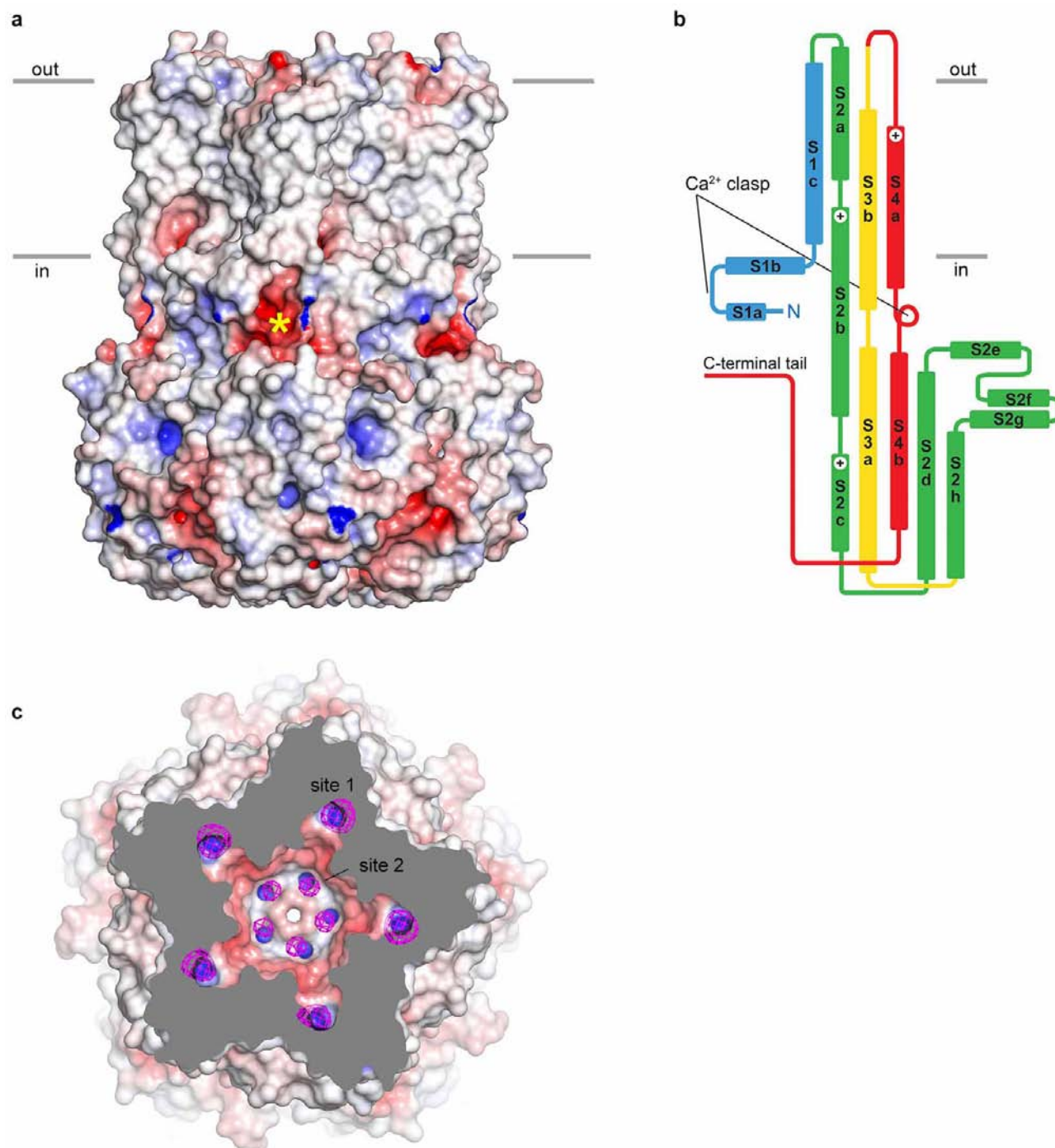
Extended Data Figure 6 | Molecular surface, subunit topology and anion binding in the outer entryway. a, The molecular surface of the channel is shown in the same orientation as Fig. 2a and coloured according to electrostatic potential (red, $-10\,kT\,e^{-1}$; grey, neutral; blue, $+10\,kT\,e^{-1}$). An asterisk marks the location of the acidic cluster in the foreground. Approximate boundaries for the membrane are indicated. **b**, Subunit topology. N-terminal ends of α-helices exposed to the pore are indicated by $+$. The colouring corresponds to that of Fig. 2b. **c**, Anion binding in the outer entryway. Extracellular cut-away view of the molecular surface of BEST1 (orthogonal representation of Fig. 4a), revealing the surface of the pore (coloured by electrostatic potential; red, $-10\,kT\,e^{-1}$; white, neutral; blue, $+10\,kT\,e^{-1}$) and anomalous difference electron density for Br^- ions (magenta mesh; 45–5 Å, non-crystallographic symmetry averaged, 8σ contour) in sites 1 and 2.
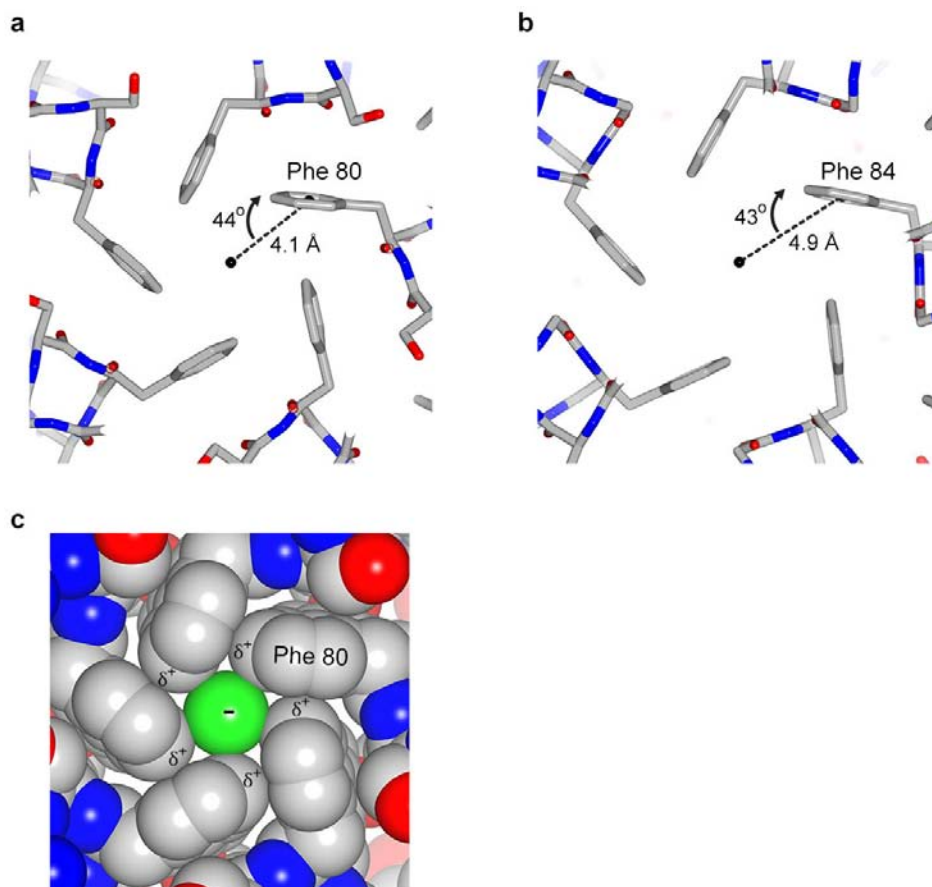
Extended Data Figure 7 | Geometry within the neck and the possibility of anion–π interactions. a, b, Representations of the pore at Phe 80 (**a**) and Phe 84 (**b**) are shown as sticks. The distance (*d*) from the central axis of the pore (black sphere) to the centre of the face of the aromatic ring is shown. An angle θ is defined as the angle between this distance vector and the plane of the ring. The geometry indicated corresponds to the crystal obtained in cymal-6. For the cymal-6-NG crystal, the values are: $d = 3.9$ Å, $\theta = 45°$ (Phe 80) and $d = 4.8$ Å, $\theta = 44°$ (Phe 84). **c,** Space-filling CPK representation of the pore at Phe 80, showing a hypothetical Cl$^-$ (green) positioned in the centre. Standard radii were used for the figure (carbon = 1.7 Å; Cl$^-$ = 1.81 Å). δ^+ and – represent partial charges on the edge of the aromatic rings and the charge on Cl$^-$, respectively.
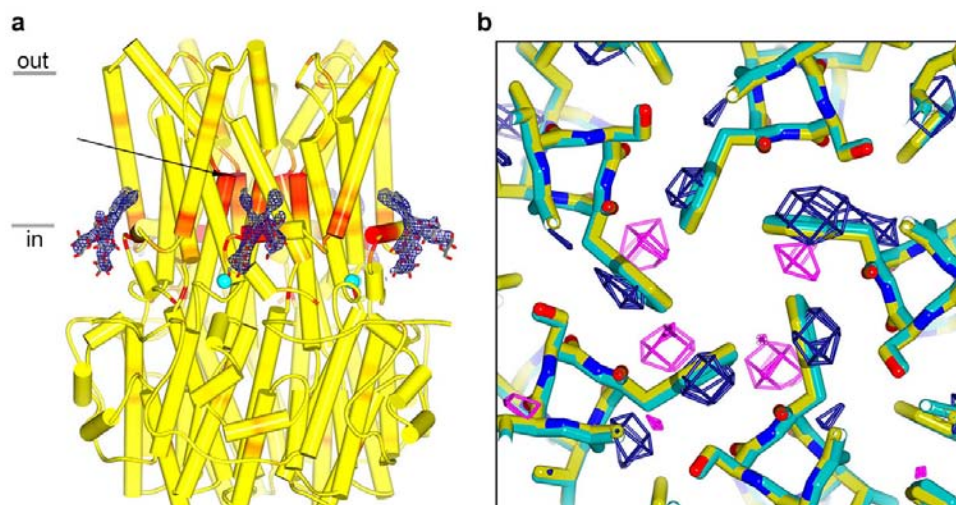
Extended Data Figure 8 | Evidence for coupling between the Ca^{2+} clasp and the gate from crystals grown in different detergents. Comparison among crystals grown using different detergents gives insight into the channel's gate and it's coupling to Ca^{2+}. Well-diffracting crystals belonging to the $P2_1$ space group were obtained using either the detergent cymal-6 or the detergent cymal-6-NG. Electron density maps indicated the presence of ordered cymal-6-NG but not cymal-6 molecules bound to the S1a–S1b components of the Ca^{2+} clasps (**a**). In addition, difference Fourier electron density maps suggested a slight widening of the neck of the pore in the structure with cymal-6 (**b**). Accordingly, while refined structures superimpose with an overall root mean squared deviation of only 0.15 Å, the diameter of the pore in the hydrophobic neck is ~0.5 Å wider at Phe 80 for crystals in cymal-6 than it is with cymal-6-NG. Differences on the order of 0.3 Å between the atomic models are localized to the region near the Ca^{2+} clasp and to the neck of the pore (**a**). The subtle effects are an indication that changes in or around the Ca^{2+} clasp induce changes in the neck of the pore and they may hint at the mechanism of gating. **a**, $2F_O - F_C$ electron density for cymal-6-NG detergent molecules, contoured at 1.2σ, is shown as blue mesh in the context of the channel. The channel, with α-helices depicted as cylinders, is coloured on a yellow-to-red spectrum according to the displacement of Cα atoms between the refined atomic models obtained from crystals grown in cymal-6 and cymal-6-NG. Yellow colour represents displacements less than 0.15 Å and red colour represents displacements greater than 0.3 Å. An arrow indicates the neck of the pore and teal spheres denote Ca^{2+}. **b**, Conformational shift in the gate. Phe 80 and surrounding residues of the refined structures from crystals in cymal-6 and cymal-6-NG are shown as sticks (coloured cyan and yellow, respectively) and viewed along the channel's axis of symmetry from the extracellular side. Superimposed on this is an $F_{cymal-6} - F_{cymal-6-NG}$ difference Fourier map, which is calculated from 25 Å to 3.5 Å resolution and contoured at −3.8σ (magenta mesh) and +3.8σ (blue mesh).

Extended Data Table 1 | Data collection, phasing and refinement statistics

	Crystal 1	Crystal 2	Crystal 3	Crystal 4	Crystal 5	Crystal 6	Crystal 7
	Native	Derivative $(Ta_6Br_{12})^{2+}$	Br	Cymal-6	+ 5 mM Ca^{2+}	Native	+ 5 mM Ca^{2+}
Data Collection	NSLS X25	NSLS X25	NSLS X25	NSLS X25	NSLS X25	APS 24-ID-C	NSLS X25
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	C2	C2
Wavelength (Å)	1.100	1.2547	0.9196	1.100	1.100	1.2543	1.700
Cell dimensions (Å):							
a	98.54	98.713	98.545	98.424	98.563	325.341	329.519
b	242.904	241.606	243.268	243.24	243.065	193.845	195.147
c	172.757	171.130	172.363	174.302	173.094	240.323	241.065
$\alpha=\gamma=90°$; $\beta= (°)$	93.68	92.478	93.71	93.29	93.65	127.22	127.09
Resolution (Å)	40 - 2.85 (2.95 - 2.85)	50 - 4.4 (4.56 - 4.4)	45 - 3.0 (3.1 - 3.0)	57 - 2.9 (3.0 - 2.9)	35 - 3.0 (3.1 - 3.0)	50 - 3.1 (3.2 - 3.1)	60 - 4.0 (4.14 - 4.0)
R_{merge}	0.113 (>1)	0.325 (>1)	0.239 (>1)	0.189 (>1)	0.242 (>1)	0.187 (>1)	0.236 (>1)
R_{pim}	0.057 (0.68)	0.106 (0.238)	0.122 (0.95)	0.091 (>1)	0.131 (>1)	0.108 (>1)	0.262 (>1)
$CC_{1/2}$ in outer shell	0.80	0.80	0.35	0.41	0.49	0.13	0.20
I/σI	14.2 (1.1)	7.5 (2.6)	7.4 (0.8)	6.9 (0.43)	7.8 (0.67)	6.8 (0.5)	2.5 (0.47)
Completeness (%)	100 (100)	99.7 (98.2)	100 (100)	99.4 (98.9)	99.4 (99.0)	99.0 (99.5)	99.9 (99.3)
Multiplicity	6.8 (6.7)	9.9 (7.6)	13.4 (12.25)	8.5 (8.1)	18.3 (18.6)	6.9 (6.9)	5.4 (5.3)
Refinement							rigid body
Resolution (Å)	40 - 2.85 (2.95 - 2.85)		45 - 3.0 (3.1 - 3.0)	57 - 2.9 (3.0 - 2.9)	35 - 3.0 (3.1 - 3.0)	50 - 3.1 (3.2 - 3.1)	60 - 4.0 (4.14 - 4.0)
No. of reflections	188162 (18682)		161270 (16047)	177017 (15887)	160053 (15157)	211664 (20620)	102001 (9644)
No. atoms	31125		31125	30780	30780	61554	61554
Ligands	400		400	55	55	110	110
Water	10		10	10	10	20	20
R_{work}	0.217 (0.361)		0.242 (0.383)	0.234 (0.433)	0.236 (0.401)	0.240 (0.376)	0.277 (0.354)
R_{free}	0.234 (0.377)		0.255 (0.388)	0.254 (0.452)	0.254 (0.415)	0.261 (0.377)	0.293 (0.384)
B-factors (Å^2)	102.3		94.7	105.9	105.9	117.1	105.70
Protein	102.0		94.5	106.0	106.0	117.2	105.80
Ligands	129.4		104.5	92.7	92.7	95.4	84.50
Water	67.5		53.9	71.3	71.3	96.5	94.30
Ramachandran (%)							
Favored	95		95	93	94	95	95
Outliers	0.4		0.3	0.6	0.5	0.3	0.3
R.m.s. deviations							
Bond lengths (Å)	0.003		0.003	0.003	0.003	0.003	0.007
Bond angles (°)	0.93		0.82	0.86	0.74	0.76	0.93

Data collection statistics are from HKL3000[48]; refinement statistics are from PHENIX[53]. $CC_{1/2}$ is defined in ref. 49. Numbers in parentheses indicate the highest resolution shells and their statistics. 5% of reflections were used for calculation of R_{free}.

LETTER

H$_2$D$^+$ observations give an age of at least one million years for a cloud core forming Sun–like stars

Sandra Brünken[1], Olli Sipilä[2,3], Edward T. Chambers[1], Jorma Harju[2], Paola Caselli[3,4], Oskar Asvany[1], Cornelia E. Honingh[1], Tomasz Kamiński[5], Karl M. Menten[5], Jürgen Stutzki[1] & Stephan Schlemmer[1]

The age of dense interstellar cloud cores, where stars and planets form, is a crucial parameter in star formation and difficult to measure. Some models predict rapid collapse[1,2], whereas others predict timescales of more than one million years (ref. 3). One possible approach to determining the age is through chemical changes as cloud contraction occurs, in particular through indirect measurements of the ratio of the two spin isomers (ortho/para) of molecular hydrogen, H$_2$, which decreases monotonically with age[4–6]. This has been done for the dense cloud core L183, for which the deuterium fractionation of diazenylium (N$_2$H$^+$) was used as a chemical clock to infer[7] that the core has contracted rapidly (on a timescale of less than 700,000 years). Among astronomically observable molecules, the spin isomers of the deuterated trihydrogen cation, ortho-H$_2$D$^+$ and para-H$_2$D$^+$, have the most direct chemical connections to H$_2$ (refs 8–12) and their abundance ratio provides a chemical clock that is sensitive to greater cloud core ages. So far this ratio has not been determined because para-H$_2$D$^+$ is very difficult to observe. The detection of its rotational ground-state line has only now become possible thanks to accurate measurements of its transition frequency in the laboratory[13], and recent progress in instrumentation technology[14,15]. Here we report observations of ortho- and para-H$_2$D$^+$ emission and absorption, respectively, from the dense cloud core hosting IRAS 16293-2422 A/B, a group of nascent solar-type stars (with ages of less than 100,000 years). Using the ortho/para ratio in conjunction with chemical models, we find that the dense core has been chemically processed for at least one million years. The apparent discrepancy with the earlier N$_2$H$^+$ work[7] arises because that chemical clock turns off sooner than the H$_2$D$^+$ clock, but both results imply that star-forming dense cores have ages of about one million years, rather than 100,000 years.

We detected the ground-state rotational transition of the para spin isomer of the deuterated trihydrogen cation (para-H$_2$D$^+$) at 1.370085 THz (ref. 13) (wavelength $\lambda = 219\ \mu$m) towards IRAS 16293-2422 A/B using the German REceiver for Astronomy at Terahertz frequencies (GREAT)[14] onboard the Stratospheric Observatory For Infrared Astronomy (SOFIA)[15]. This line has so far only been tentatively detected in absorption against the bright high-mass star-forming region Orion Irc2 by the Kuiper Airborne Observatory[16]. We also observed the ground-state line of ortho-H$_2$D$^+$ at 372.421 GHz (ref. 17) ($\lambda = 0.8$ mm) towards the same source using the Atacama Pathfinder EXperiment (APEX)[18] submillimetre telescope located in the Chilean Atacama desert at an altitude of 5,100 m. IRAS 16293-2422 A/B consists of a triple system of young ($<$100,000 years) solar-type protostars, comprising a close protobinary (A1/A2) and a third protostar (B) about 600 astronomical units (AU) away from these[19,20], surrounded by a massive envelope (a dense core of about two solar masses) with steeply decreasing (from the inside outward) temperature and density distributions[21,22]. This dense core still bears the physical characteristics typical of starless cores on the verge of star formation (the so-called pre-stellar cores[23]), with the bulk of the material still at low temperatures ($T < 20$ K) and densities (number density of

H$_2$ molecules $n(\text{H}_2) < 10^6\ \text{cm}^{-3}$). The dense core is embedded in the dark cloud Lynds 1689N in Ophiuchus at a distance of 120 pc (ref. 24).

The present observations provide the measurement of the ortho/para H$_2$D$^+$ ratio, and thus the corresponding ortho/para H$_2$ ratio across the dense core (see Methods). The para-H$_2$D$^+$ spectrum observed with SOFIA and the ortho-H$_2$D$^+$ spectrum observed with APEX are shown in Fig. 1, together with the model predictions (detailed below). We note that para-H$_2$D$^+$ shows a strong and narrow absorption profile against the far-infrared continuum emission caused by the central protostellar heating of the surrounding dust grains, while ortho-H$_2$D$^+$ is observed in emission with a similar width. These observational facts are related to the cold temperatures of the environment, where almost all para-H$_2$D$^+$ is in its rotational ground state (0_{00}) and is therefore observed in absorption (see the energy level diagram in Fig. 1). Owing to the nuclear spin conversion discussed in the Methods, the ground (1_{11}) and the first excited rotational state (1_{10}) of ortho-H$_2$D$^+$ are populated even at the low temperature of the dense core. As a consequence, in combination with the lower continuum brightness at larger wavelengths, the major contribution to the ortho-H$_2$D$^+$ signal observed at 372 GHz is due to emission (see the energy level diagram in Fig. 1). In what follows, we estimate the amounts of para-H$_2$D$^+$ and ortho-H$_2$D$^+$ in the line of sight causing the observed absorption and emission features by radiative transfer modelling. It turns out that the ortho/para ratio of H$_2$D$^+$ is below 0.1 in the cool outer part of the dense core where the lines originate. This implies a very low ortho/para H$_2$ ratio in this region.

We model the observed lines using the previously derived dense core structure[22] in conjunction with chemistry and radiative transfer calculations[11,25,26] (see Methods). The radial density distribution of the dense core is described by a power law between the central cavity (radial distance 30 AU from the centre) and the outer edge of the core (6,100 AU). The gas temperature increases strongly inwards in this model owing to gas compression in the collapse and to radiation from the protostars. In agreement with several previous studies of this region[27], we assume that the dense core is embedded in an ambient cloud with typical dark cloud conditions ($n(\text{H}_2) \approx 10^4\ \text{cm}^{-3}$, $T \approx 10$ K; see Extended Data Fig. 1). According to our modelling results, most of the para-H$_2$D$^+$ absorption (83%) and nearly all ortho-H$_2$D$^+$ emission (91%) originate in the dense core at radial distances from the centre between 2,000 AU and 6,100 AU, where the kinetic temperature decreases from $\sim$20 K to $\sim$13 K, and the hydrogen density $n(\text{H}_2)$ decreases from about $10^6\ \text{cm}^{-3}$ to about $10^5\ \text{cm}^{-3}$ (see Extended Data Fig. 1). This region still preserves the conditions of the original pre-stellar core. The para- and ortho-H$_2$D$^+$ spectra produced by our best-fit model are displayed in Fig. 1, together with the observed spectra. The best-fit model predicts an average ortho/para H$_2$D$^+$ ratio of 0.07 ± 0.03 between 2,000 AU and 6,100 AU.

Such a low value for the ortho/para H$_2$D$^+$ ratio can only be understood as a temporal decrease in parallel with a decreasing ortho/para H$_2$ ratio. The time evolution of the chemical abundances in different parts of the dense core and in the ambient cloud was calculated using

[1]I. Physikalisches Institut, Universität zu Köln, Zülpicher Straße 77, 50937 Köln, Germany. [2]Department of Physics, PO Box 64, 00014 University of Helsinki, Finland. [3]Max-Planck Institut für Extraterrestrische Physik, Gießenbachstraße 1, 85741 Garching bei München, Germany. [4]School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK. [5]Max-Planck Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany.
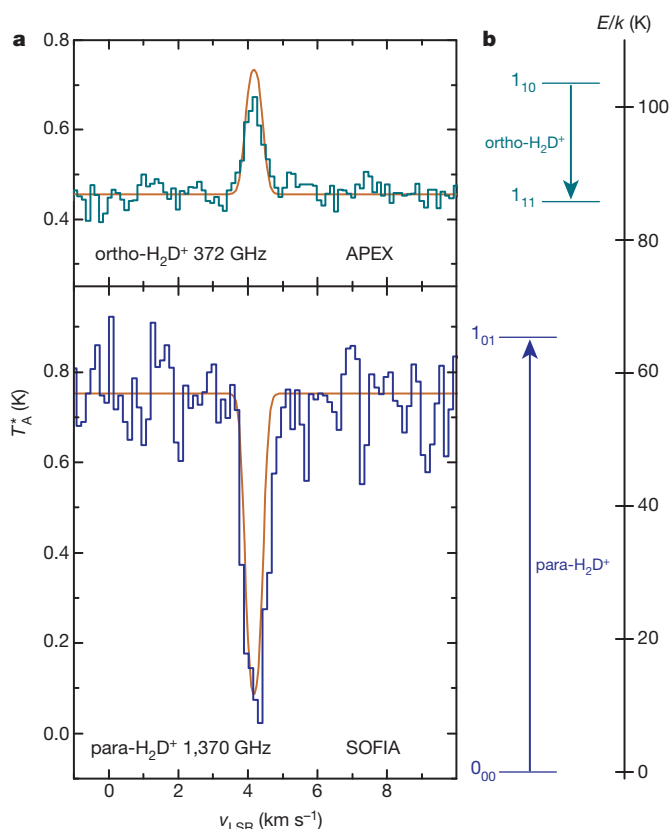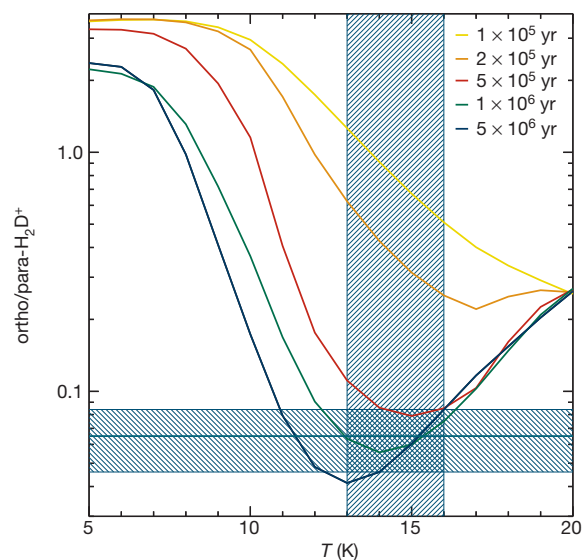
Figure 1 | Observed and modelled H_2D^+ spectra. a, The histograms show the ortho-H_2D^+ (top) and para-H_2D^+ (bottom) rotational ground-state lines as observed with APEX/FLASH and SOFIA/GREAT, respectively; the orange lines show the modelled line profiles. Intensities are given as antenna temperatures T_A^* and v_{LSR} denotes the velocity with respect to the local standard of rest. **b**, Energy level diagram (in units of temperature, E/k, where k is the Boltzmann constant) of the lowest rotational states of ortho- and para-H_2D^+.

Figure 2 | Modelled ortho/para H_2D^+ abundance ratio. At kinetic temperatures T above ~ 12 K, the ortho/para H_2D^+ ratio is completely determined in reactions with ortho- and para-H_2, and it is closely tied to the evolution of the ortho/para H_2 ratio. The shaded vertical region indicates the temperature range applicable to the dense core surrounding IRAS 16293-2422 A/B (at radial distances from the core centre of 3,000–6100 AU), while the horizontal shade indicates the observed ortho/para H_2D^+ ratio. Together, these limits suggest a dense core age of at least one million years. The gas density, $n(H_2) = 10^5$ cm^{-3}, and the visual extinction, $A_V = 10$ mag, are kept constant in this model.

our gas-grain chemistry model[25]. The resulting radial abundance distributions of para- and ortho-H_2D^+, together with the density, temperature and velocity profiles, were used as input for a Monte Carlo radiative transfer program[26] designed to predict observable line profiles. The excitation of the rotational transitions of H_2D^+ in collisions with para- and ortho-H_2 are calculated using theoretical state-to-state rate coefficients[11]. The slow conversion of ortho- to para-H_2, together with the coupling of the ortho/para H_2 ratio to that of H_3^+ and its deuterated species through proton exchange reactions (see Extended Data Fig. 2) allows us to use the observed ortho/para H_2D^+ ratio as a chemical clock for the dense core age since the time of its formation within the ambient cloud.

Using conservative values of the initial ortho/para H_2 ratio in our time-dependent chemical models (see Methods), the low values of the ortho/para H_2D^+ ratio ($\sim 0.065 \pm 0.019$) found in the outermost layers of the dense core with $T = 13$–16 K can only be reached after about a million years of chemical evolution, preceded by a period at least equally long in conditions corresponding to the embedding ambient dark cloud. To illustrate the temporal evolution of the ortho/para H_2D^+ ratio in conditions corresponding to the dense core surrounding IRAS 16293-2422 A/B, we plot this ratio in Fig. 2 as a function of the kinetic temperature of the environment for different evolution times after the formation of the dense core. Owing to the restrictions in dense core temperature and the observed ortho/para H_2D^+ ratio (shown as vertical and horizontal shaded areas, respectively, in Fig. 2), the temporal evolution of the dense core is at least one million years (see Methods).

Therefore, we have verified that the observed ortho/para H_2D^+ ratio is setting limits on the core age. The ortho/para H_2D^+ ratio gives a more

direct estimate of the ortho/para H_2 ratio than the previously used deuterium fraction measurement of N_2H^+ (that is, N_2D^+/N_2H^+; ref. 7), in particular for evolved regions with ortho/para H_2 ratios of less than 0.01. Below this value, the N_2D^+/N_2H^+ ratio loses correlation with the ortho/para H_2 ratio (see Extended Data Figs 3 and 4). Therefore, at this point, the N_2D^+/N_2H^+ chemical clock stops while the clock based on the ortho/para H_2D^+ ratio keeps running. Our results indicate that the average ortho/para H_2 ratio is about 2×10^{-4} between radii of 3,000 AU and 6,100 AU ($T = 13$–16 K), which can be reproduced only at very late times of chemical evolution (see Extended Data Fig. 5). Our conservative analysis gives an age estimate of at least one million years. The very low value of ortho/para H_2 found in the core around IRAS16293-2422 is hardly possible to probe by any other means, and we conclude that the timing set by the ortho/para H_2D^+ ratio is most relevant for constraining the duration of the dense cloud core phase in the course of star formation.

1. Ward-Thompson, D. *et al.* An observational perspective of low-mass dense cores. II: Evolution toward the initial mass function. In *Protostars and Planets V* (eds Reipurth, B., Jewitt, A. P. & Keil, K.) 33–46 (Univ. Arizona Press, 2007).
2. Hartmann, L., Ballesteros-Paredes, J. & Heitsch, F. Rapid star formation and global gravitational collapse. *Mon. Not. R. Astron. Soc.* **420,** 1457–1461 (2012).
3. Mouschovias, T. Ch., Tassis, K. & Kunz, M. W. Observational constraints on the ages of molecular clouds and the star formation timescale: ambipolar-diffusion-controlled or turbulence-induced star formation? *Astrophys. J.* **646,** 1043 (2006).
4. Flower, D. R. & Watt, G. D. On the ortho-H_2/para-H_2 ratio in molecular clouds. *Mon. Not. R. Astron. Soc.* **209,** 25–31 (1984).
5. Flower, D. R., Pineau Des Forêts, G. & Walmsley, C. M. The importance of the ortho:para H_2 ratio for the deuteration of molecules during pre-protostellar collapse. *Astron. Astrophys.* **449,** 621–629 (2006).
6. Pagani, L., Roueff, E. & Lesaffre, P. Ortho-H_2 and the age of interstellar dark clouds. *Astrophys. J.* **739,** L35 (2011).

7. Pagani, L. *et al.* Ortho-H_2 and the age of prestellar cores. *Astron. Astrophys.* **551,** A38 (2013).
8. Pagani, L., Salez, M. & Wennier, P. G. The chemistry of H_2D^+ in cold clouds. *Astron. Astrophys.* **258,** 479–488 (1992).
9. Gerlich, D., Herbst, E. & Roueff, E. $H_3^+ + HD \rightarrow H_2D^+ + H_2$: low-temperature laboratory measurements and interstellar implications. *Planet. Space Sci.* **50,** 1275–1285 (2002).
10. Hugo, E., Asvany, O., Harju, J. & Schlemmer, S. Toward understanding of H_3^+ isotopic and nuclear spin fractionations in cold space. In *Molecules in Space and Laboratory* (eds Lemaire, J. L. & Combes, F.) 119 (S. Diana, 2007).
11. Hugo, E., Asvany, O. & Schlemmer, S. $H_3^+ + H_2$ isotopic system at low temperatures: microcanonical model and experimental study. *J. Chem. Phys.* **130,** 164302 (2009).
12. Sipilä, O. *et al.* Modelling line emision of deuterated H_3^+ from prestellar cores. *Astron. Astrophys.* **509,** A98 (2010).
13. Asvany, O. *et al.* High-resolution rotational spectroscopy in a cold ion trap: H_2D^+ and D_2H^+. *Phys. Rev. Lett.* **100,** 233004 (2008).
14. Heyminck, S. *et al.* GREAT: the SOFIA high-frequency heterodyne instrument. *Astron. Astrophys.* **542,** L1 (2012).
15. Young, E. T. *et al.* Early science with SOFIA, the Stratospheric Observatory For Infrared Astronomy. *Astrophys. J.* **749,** L17 (2012).
16. Boreiko, R. T. & Betz, A. L. A search for the rotational transitions of H_2D^+ at 1370 GHz and H_3O^+ at 985 GHz. *Astrophys. J.* **405,** L39–L42 (1993).
17. Amano, T. & Hirao, T. Accurate rest frequencies of submillimeter-wave lines of H_2D^+ and D_2H^+. *J. Mol. Spec.* **233,** 7 (2005).
18. Güsten, R. *et al.* The Atacama Pathfinder EXperiment (APEX)—a new submillimeter facility for southern skies. *Astron. Astrophys.* **454,** L13 (2006).
19. Wootten, A. The duplicity of IRAS 16293–2422: a protobinary star? *Astrophys. J.* **337,** 858–864 (1989).
20. Girart, J. M., Estalella, R., Palau, A., Torrelles, J. M. & Rao, R. On the origin of the molecular outflows in IRAS 16293–2422. *Astrophys. J.* **780,** L11 (2014).
21. Stark, R. *et al.* Probing the early stages of low-mass star formation in LDN 1689N: dust and water in IRAS 16293–2422A, B, and E. *Astrophys. J.* **608,** 341–364 (2004).
22. Crimier, N. *et al.* The solar type protostar IRAS16293–2422: new constraints on the physical structure. *Astron. Astrophys.* **519,** A65 (2010).
23. Keto, E. & Caselli, P. Dynamics and depletion in thermally supercritical starless cores. *Mon. Not. R. Astron. Soc.* **402,** 1625 (2010).
24. Lombardi, M., Lada, C. J. & Alves, J. Hipparcos distance estimates of the Ophiuchus and the Lupus cloud complexes. *Astron. Astrophys.* **480,** 785–792 (2008).
25. Sipilä, O., Caselli, P. & Harju, J. HD depletion in starless cores. *Astron. Astrophys.* **554,** A92 (2013).
26. Juvela, M. Non-LTE radiative transfer in clumpy molecular clouds. *Astron. Astrophys.* **322,** 943–961 (1997).
27. Le Gal, R. *et al.* Interstellar chemistry of nitrogen hydrides in dark clouds. *Astron. Astrophys.* **562,** A83 (2014).

Author Contributions S.S., S.B., O.A., P.C., J.H., O.S. and J.S. jointly designed the study and proposed the SOFIA observations. E.T.C. performed the calibration and the analysis of the SOFIA data. C.E.H. was instrumental in developing the GREAT receiver. T.K. and K.M.M. made the APEX observations and analysed these data. O.S. carried out the chemistry and radiative transfer modelling with help from J.H. The paper was jointly written by S.B., J.H., O.S., P.C. and S.S. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.B. (bruenken@ph1.uni-koeln.de) or S.S. (schlemmer@ph1.uni-koeln.de).

METHODS

Observational strategy. Although the rotational ground-state emission line of ortho-H_2D^+ has been observed towards several cold starless and low-mass star-forming cores[21,28], para-H_2D^+ has previously been detected only tentatively, in absorption towards the Orion IRc2 region using the NASA Kuiper Airborne Observatory (KAO)[16]. The ground-state rotational transition of para-H_2D^+ at 1.37 THz is extremely difficult to observe from the ground owing to poor atmospheric transmission and the resulting large system temperatures at terahertz frequencies. The transition frequency was not covered by the Herschel/HIFI bands[29]. The GREAT[14] receiver onboard SOFIA[15] now enables us to observe para-H_2D^+. Despite its rather high upper-state energy (66 K above ground) the para-H_2D^+ line at 1.37 THz can be excited in cold, dense cores. However, the expected brightness temperature of the emission line is very weak. Absorption against a bright continuum source is the best way of detecting the line with currently available instruments in a reasonable observing time. Because H_2D^+ is most abundant in cold gas deficient in CO[30], the best chance of detecting the para-H_2D^+ line is by observing it towards a young (Class 0)[31] protostar with bright continuum emission surrounded by a massive, cool envelope. IRAS 16293-2422 A/B is one of the best targets fulfilling these criteria.

SOFIA observations. We observed the para-H_2D^+ ground-state transition $J_{KaKc} = 0_{00}-1_{01}$ at 1370.085 GHz ($\lambda = 219 \mu$m)[13] towards IRAS 16293-2422 A/B using the GREAT instrument[14] onboard SOFIA[15] on 23 July 2013 during SOFIA's southern deployment to New Zealand in Cycle 1 operations. The target position was: right ascension 16 h 32 m 22.9 s, declination $-24°28'39''$ (J2000). We used the GREAT L1a channel (1.25–1.40 THz) with the XFFTS backend[32], which has a bandwidth of 2.5 GHz and a channel spacing of 88 kHz (17 ms^{-1}). We operated the instrument in the double beam-switch mode, with a chopping frequency of 1 Hz, a chop amplitude of 20'' (for a total beam throw of 40''), and a chop angle of 0° from horizontal in the telescope reference frame. In total, we had an on-source integration time of ~26 min (system temperature $T_{sys} = 1,760$ K). We calibrated the data using the standard pipeline (*kalibrate*[33]), which fits an atmospheric model to the observed sky to calculate the atmospheric opacity. The forward and main beam efficiencies for the L1 channel on GREAT are 0.97 and 0.67, respectively. The half-power beam width (HPBW) of the 2.5-m SOFIA telescope is ~22'' at 1,370 GHz.

During observations we noticed a slow drop-off of the continuum level as the telescope drifted away from the nominal source position, caused by the lack of good tracking stars in the heavily extinct region around IRAS 16293-2422 A/B. We had to re-acquire the source three times. Because the continuum level is of vital importance for this absorption measurement, we averaged our data in such a way that the nominal continuum value of the source was conserved. To determine this continuum level, we averaged the continuum level calculated in each of the spectra obtained after reacquisition (four chop-nod pairs) using a large number of line-free channels. The nominal continuum intensity, measured on the equivalent antenna temperature scale, of $T_{A,C}^* = 0.79 \pm 0.03$ K, obtained in this way agrees very well with the flux density of 460 Jy that we obtained from an analysis of archival Herschel/SPIRE (Spectral and Photometric Imaging Receiver) and PACS (Photoconductor Array Camera and Spectrometer) continuum maps. This nominal value was then used to re-scale all of the other averaged pairs. We calculated the weighted average of the remaining spectra using a $1/\sigma_{rms}^2$ weighting, so that the spectra with the weakest original continuum (the highest root-mean-square (r.m.s.) noise after re-scaling) have the smallest contribution to the final spectrum. Finally, to correct for the double-sideband reception of GREAT, we subtracted half of the continuum from the spectrum (assuming equal gains in the two sidebands). A Gaussian fit to the para-H_2D^+ absorption spectrum smoothed to a velocity resolution of 0.13 km s^{-1} gives the following line parameters: velocity with respect to the local standard of rest $v_{LSR} = 4.24 \pm 0.02$ km s^{-1}, line width (in units of velocity) $\Delta v = 0.73 \pm 0.05$ km s^{-1}, and $T_A^* - T_{A,C}^* = -0.70 \pm 0.07$ K (the difference between the 219-μm continuum and the antenna temperature at the line centre). The observed para-H_2D^+ spectrum is compatible with total absorption in the line centre, which implies a very large optical depth and a low excitation temperature for the para-H_2D^+ line.

APEX observations. The target position as indicated above was observed in the $J_{KaKc} = 1_{10}-1_{11}$ ortho ground-state transition of H_2D^+ at 372.421 GHz ($\lambda = 805 \mu$m)[17] using the 12-m APEX telescope[18] on 5 and 14 August 2013 in excellent weather conditions. We used the lower-frequency module (covering 262–374 GHz) of the upgraded version of the First Light APEX Submillimetre Heterodyne instrument (FLASH[34]). FLASH is a two-sideband receiver with a bandwidth of 4 GHz for each sideband. On the two observing days we used different frequency settings that covered the sky frequency of the ortho-H_2D^+ line. We employed the 4-GHz total bandwidth (per intermediate frequency) of the newest version of the APEX facility Fast Fourier Transform Spectrometer (FFTS[35]). The FFTS band was split into 104,859 channels with a spacing of 38.2 kHz, which corresponds to 31 m s^{-1} at 372.4 GHz. The calibration was achieved by the standard chopper-wheel method. The background was subtracted by means of wobbling sub-reflector, using a beam throw of 150'' and a switching rate of 0.7 Hz. In total we spent ~50 min on source
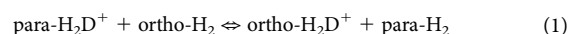
($T_{sys} = 590$ K). Conversion of the measured antenna temperatures to flux density units (in Jy) and a main-beam brightness temperature, T_{MB}, scale (in K) was established by interpolating previously determined aperture and main-beam efficiencies[18], η_A and η_{MB}, which yielded $\eta_A = 0.58$ and $\eta_{MB} = 0.68$. The HPBW of the antenna is 17'' at 372 GHz. Fitting a Gaussian to our pointing drift scans yields a deconvolved source size of 12'' (FWHM). The baseline level in the spectra gave a flux density of 18.6 ± 1.7 Jy for the continuum source (antenna temperature at APEX $T_{A,C}^* = 0.47 \pm 0.04$ K). These values are in good agreement with previous measurements[21]. We estimate a 20% error for our intensity calibration from the difference of the two spectra summed up for each observing date. This is larger than typical FLASH calibration errors, but explainable by the fact that the line frequency lies in the wings of deep atmospheric O_2 and H_2O absorption lines (at 368 GHz and 380 GHz, respectively). A Gaussian fit to the ortho-H_2D^+ spectrum smoothed to a velocity resolution of 0.12 km s^{-1} gives the following line parameters: $v_{LSR} = 4.17 \pm 0.02$ km s^{-1}, $\Delta v = 0.62 \pm 0.05$ km s^{-1}, and $T_A^* - T_{A,C}^* = 0.21 \pm 0.02$ K, consistent with the velocity position and width of the para-H_2D^+ line at 1.37 THz.
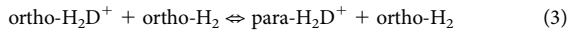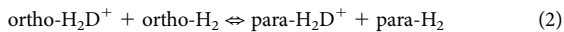
Both our peak and our integrated main-beam brightness temperatures are a factor of ~3 lower than the published values measured with the 15-m James Clerk Maxwell Telescope (JCMT)[21]. These previous measurements were probably positioned within a few arcseconds of our pointing. The large difference is not explainable by the slightly different antenna sizes. We found archival H_2D^+ spectra towards our target source taken with the HARP instrument on the JCMT on three different dates, 7 and 8 August 2007, and 21 February 2008. The summed spectrum, while noisy, agrees with ours within the uncertainties.

Source model. The dense core surrounding IRAS 16293-2422 A/B is known to have steep density and temperature gradients. Therefore, the standard method for deriving column densities from observed spectra based on the assumption of line-of-sight homogeneity is not likely to give reliable results. We adopt a frequently used physical model for IRAS 16293-2422 A/B (ref. 22), where the radial density distribution of the dense core is described by a power law, $n(H_2) \propto r^{-1.8}$, between the central cavity (30 AU) and the outer edge (6,100 AU, corresponding to 51'' at a distance of 120 pc), see Extended Data Fig. 1. The gas temperature increases strongly inwards in this model because of gas compression in the collapse and due to the radiation from the protostars. Infall speeds are significant in warm inner parts of the dense core where the abundance of H_2D^+ is negligible. Recent observations with Herschel/HIFI[27,36] imply the presence of a low-density absorbing layer in front of the dense core, which can probably be attributed to the ambient dark cloud Lynds 1689. We therefore add to the dense core a spherically symmetric, ambient cloud ($n(H_2) = 10^4$ cm^{-3}, $T = 10$ K) with a thickness causing a visual extinction of $A_V = 10$ mag to the outer edge of the dense cloud. Our radiative transfer calculations show that this ambient cloud deepens slightly the para-H_2D^+ absorption, and causes self-absorption to the ortho-H_2D^+ emission. For the purposes of chemistry modelling and radiative transfer calculations, the model is divided into concentric shells where the density and temperature are assumed to be constant.

The ortho/para H_2 ratio. Molecular hydrogen is formed when two hydrogen atoms react on dust grains. The spins of the two protons ($I = 1/2$) in H_2 give rise to four nuclear spin states, three with total nuclear spin $I = 1$ (ortho-H_2) and degenerate spin orientations ($m_I = -1, 0, 1$), and one with total nuclear spin $I = 0$ (para-H_2) and no degeneracy ($m_I = 0$). Each of those states is formed with equal probability, implying a statistical ortho/para ratio of H_2, of 3:1. As a result of the Pauli exclusion principle, ortho nuclear spin states are connected to the energy states with odd rotational quantum numbers J, while para spin states are found at rotational levels with even J. As a consequence, the odd (ortho) and even (para) rotational state populations of H_2 are far from thermodynamical equilibrium upon formation, especially in cold molecular clouds. After entering the gas phase, the ortho/para-H_2 ratio is altered by proton-exchange reactions, whereas conversion between ortho- and para-H_2 by radiation and inelastic collisions is spin-forbidden. The dominant spin-changing reactions are those with H^+ and H_3^+. These well-studied reactions[9,11,37,38] can thermalize ortho/para-H_2 efficiently in warm gas, whereas below 20 K, ortho/para-H_2 approaches thermal equilibrium very slowly as proton exchange reactions have to compete with the more favoured ortho production on grains. At the typical dark cloud kinetic temperature, $T = 10$ K, the thermal ortho/para-H_2 is as low as 3.6×10^{-7}. However, this value is probably never reached. According to chemistry models[5,25,39], ortho/para-H_2 remains suprathermal in very cold gas (<12 K). It is the subtle detail of the indistinguishable two fermions (protons) in H_2, in combination with the spin changing proton exchange reactions, that turn ortho/para-H_2 into a robust chemical clock in cold molecular clouds[4,5,39] (see also Extended Data Fig. 5).

Analytical relation between the H_2 and H_2D^+ ortho/para ratios. The ortho/para ratio of H_2D^+ in molecular clouds is mainly regulated by the following chemical reactions[9,10]:

$$\text{para-}H_2D^+ + \text{ortho-}H_2 \Leftrightarrow \text{ortho-}H_2D^+ + \text{para-}H_2 \qquad (1)$$

$$\text{ortho-H}_2\text{D}^+ + \text{ortho-H}_2 \Leftrightarrow \text{para-H}_2\text{D}^+ + \text{para-H}_2 \qquad (2)$$

$$\text{ortho-H}_2\text{D}^+ + \text{ortho-H}_2 \Leftrightarrow \text{para-H}_2\text{D}^+ + \text{ortho-H}_2 \qquad (3)$$

The reaction ortho-H_2D^+ + ortho-$\text{H}_2 \rightarrow$ ortho/para-H_3^+ + HD, returning deuterium back to deuterated hydrogen (HD), occurs about three times more slowly than the ortho-to-para conversion described by reactions (2) and (3) (ref. 11). Also, the further conversion of H_2D^+ to D_2H^+ is slower than the ortho-to-para conversion because ortho-H_2 is always more abundant than HD. From the reaction system described above, the following analytic expression can be derived for the equilibrium ortho/para-H_2D^+ (ref. 10):

$$\frac{[\text{ortho-H}_2\text{D}^+]}{[\text{para-H}_2\text{D}^+]} = \frac{(k_1^+ + k_3^-) \times [\text{ortho-H}_2]/[\text{para-H}_2] + k_2^-}{(k_2^+ + k_3^+) \times [\text{ortho-H}_2]/[\text{para-H}_2] + k_1^-}$$

where $k_1^+, k_1^-, k_2^+, k_2^-, k_3^+$ and k_3^- are the rate coefficients of the forward (+) and backward (−) direction of reactions (1), (2) and (3). As shown below, this simple relation, using the Arrhenius behaviour of the rate coefficients[9,10], approximates the ortho/para-H_2D^+ predicted by comprehensive chemical models. This emphasizes the direct correlation between the ortho/para ratios of H_2D^+ and H_2 that is the central tool of this work.

Chemistry model. The time-evolution of the chemical abundances in different parts of the dense core surrounding IRAS 16293-2422 A/B and in the embedding ambient cloud is calculated using a chemistry model containing reaction sets for both gas-phase and grain surface chemistry[25]. The gas-phase reaction set is based on the publicly available Ohio State University reaction set (available upon request from Eric Herbst, eh2ef@virginia.edu), which has been expanded to include the spin states of the light hydrogen-bearing species H_2^+, H_2, and H_3^+. In addition, deuterated species with up to four atoms and their reaction rates are included. Similar ortho/para separation and deuteration as in the gas phase is applied to the surface reaction set, which is based on a previously published model[40]. The model is pseudo-time-dependent, that is, we follow the chemical evolution assuming that the dense core and the ambient cloud are static. We assume that the initial chemical composition of the dense core is determined in conditions corresponding to the ambient cloud ($n(\text{H}_2) = 10^4$ cm^{-3}). Therefore, we calculate molecular abundances at different times in the ambient cloud and use these abundances as initial conditions for the dense core model. In the ambient cloud, the gas is assumed to be initially atomic, with the exception of hydrogen, which is molecular. We have fixed the cosmic-ray ionization rate to $\zeta = 1.3 \times 10^{-17}$ s^{-1}, the grain radius to $a_g = 0.1$ μm, and the initial ortho/para-H_2 to 1.0×10^{-3}, corresponding to a spin temperature of $T_{\text{spin}} \approx 20$ K. This assumption is based on the fact that the ortho/para ratio is possibly thermalized by collisions with protons (H^+) in warm gas down to ~20 K during the contraction and cooling phase of the cloud. Efficient thermalization down to ~30 K has been demonstrated previously[5,39,41]. Starting the simulation from an initial ortho/para-H_2 of 0.5 ($T_{\text{spin}} \approx 60$ K), which is typical for diffuse interstellar clouds[42], it takes about 3–4 million years of chemical processing in conditions corresponding to the ambient cloud to reduce ortho/para-H_2 to ~10^{-3}. By comparison with the observed ortho- and para-H_2D^+ lines, we obtain the same chemical age of about a million years for the dense core (see below) using either the high and the low initial ortho/para-H_2 ratio, although in the former case the ambient cloud has to be very old (more than 3 million years).

The ortho/para-H_2D^+ ratio, as a function of ortho/para-H_2, resulting from the full simulation closely follows the analytical formula presented above[10]. This is illustrated in Extended Data Fig. 2, which shows the relationship for different kinetic temperatures and for ortho/para-$\text{H}_2 < 0.1$, roughly corresponding to times after 100,000 years of chemical evolution. The fact that reactions (1) to (3) dominate the relative abundances of ortho-H_2D^+ and para-H_2D^+ in this regime can also be verified by inspecting the actual reaction rates during the simulation. At low values of ortho/para-H_2 for which $(k_2^+ + k_3^+) \times$ ortho/para-$\text{H}_2 \ll k_1^-$, ortho/para-H_2D^+ is linearly proportional to ortho/para-H_2. For high values of ortho/para-H_2 for which $(k_2^+ + k_3^+) \times$ ortho/para-$\text{H}_2 \gg k_1^-$, ortho/para-H_2D^+ approaches a constant, which, according to the analytical formula, is determined by the ratio $(k_1^+ + k_3^-)/(k_2^+ + k_3^+)$. The full simulation predicts a slightly lower asymptotic value of ortho/para-H_2D^+, which can be attributed mainly to the reaction ortho-H_2D^+ + ortho-$\text{H}_2 \rightarrow$ ortho/para-H_3^+ + HD.

The disagreement between the analytical formula and the simulation is most marked at very low temperatures (<10 K), where the deuteration reactions H_3^+ + HD $\rightarrow \text{H}_2\text{D}^+$ + H_2 and H_2D^+ + HD $\rightarrow \text{D}_2\text{H}^+$ + H_2 influence ortho/para-H_2D^+. The importance of these reactions depends on the H_3^+ abundance, which is sensitive to the cosmic-ray ionization rate ζ and the grain radius a_g. When $T > 10$ K, the relationship between ortho/para-H_2D^+ and ortho/para-H_2 is almost independent of other physical conditions than the kinetic temperature. The ortho/para-H_2 ratio has been previously derived by modelling the deuterium fraction of the HCO^+ and

N_2H^+ molecular ions[6,7,39]. This is based on the fact that the $\text{DCO}^+/\text{HCO}^+$ and $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+$ abundance ratios show a general increasing trend with a decreasing ortho/para-H_2. As shown in Extended Data Figs 3 and 4, the relationships $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+$ and $\text{DCO}^+/\text{HCO}^+$ (not plotted) versus ortho/para-H_2 also depend (besides on T) on $n(\text{H}_2)$, ζ, and a_g. The ortho/para-H_2D^+ versus ortho/para-H_2 relationship for the same parameter space is shown for comparison. It is evident that in conditions where ortho/para-$\text{H}_2 < 0.01$ (corresponding roughly to $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+ > 0.01$), ortho/para-$\text{H}_2\text{D}^+$ gives a better estimate for ortho/para-H_2 than $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+$. We note, however, that the accuracy of the $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+$ method can be substantially improved by mapping observations of the N_2D^+ and N_2H^+ distributions[7]. At even longer evolutionary times $\text{N}_2\text{D}^+/\text{N}_2\text{H}^+$ actually becomes independent of ortho/para-H_2, whereas ortho/para-H_2D^+ can still be used as a chemical clock. This is the regime of ortho/para-H_2 that is relevant in studies of old pre-stellar cores.

Radiative transfer calculations. The resulting radial abundance distributions of para-H_2D^+ and ortho-H_2D^+, together with the density, temperature and velocity profiles, were used as input for a Monte Carlo radiative transfer program[26] to predict the line profiles observed with the telescopes used in the present study. The excitation of the rotational transitions of H_2D^+ in collisions with para- and ortho-H_2 are calculated using theoretically determined state-to-state rate coefficients[11]. We calculate the radial distributions of the optical thicknesses and the excitation temperatures of the ground-state transitions of para- and ortho-H_2D^+ as a function of velocity, and construct the observable absorption/emission spectra, taking into account the continuum source in the centre of the dense core, and the beam profile of the respective telescopes. We ran multiple models corresponding to different ages of the initial cloud and of the dense core itself. Six different time steps between 10,000 years and 2 million years were considered. We searched for the best match between the modelled and observed line profiles by performing a χ^2 test for each model, simultaneously for the two lines. From this analysis, we find that the model with a million years of chemical evolution in both the initial cloud stage and in the dense core yields the best fit to the observations. In this case the optical thickness in the line centre is determined to be $\tau_o = 0.33$ for the ortho-H_2D^+ line and to be $\tau_p = 2.7$ for the para-H_2D^+ line. The fractional abundance relative to H_2 of both ortho- and para-H_2D^+ increases with the distance from the source centre, reaching 10^{-10} and 10^{-9}, respectively, at the outer edge of the envelope (at a radius of 6,100 AU).

Possibility of detecting para-H_2D^+ in other sources. IRAS 16293-2422 A/B is one of the brightest far-infrared sources in nearby molecular clouds and provides a particularly favourable target for observing para-H_2D^+ in absorption. A quick look at archival Herschel continuum maps of nearby complexes[43], including Chamaeleon, Corona Australis, Ophiuchus, Perseus, Serpens and Taurus, reveals eight embedded Class 0/I protostars or protoclusters with far-infrared flux densities at least 25% of that of IRAS 16293. We estimate that para-H_2D^+ absorption from a dense core similar to that surrounding IRAS 16293 could be detected towards these weaker sources in approximately 1.5 h with SOFIA/GREAT. The Herschel maps, together with Spitzer archival catalogues, can be used to select embedded sources with massive envelopes that are likely to be most appropriate for para-H_2D^+ absorption observations[44].

28. Caselli, P. et al. Survey of ortho-H_2D^+ ($1_{1,0}$-$1_{1,1}$) in dense cloud cores. *Astron. Astrophys.* **492,** 703–718 (2008).
29. de Graauw, T. et al. The *Herschel*-Heterodyne Instrument for the Far-Infrared (HIFI). *Astron. Astrophys.* **518,** L6 (2010).
30. Caselli, P., van der Tak, F. F. S., Ceccarelli, C. & Bacmann, A. Abundant H_2D^+ in the pre-stellar core L1544. *Astron. Astrophys.* **403,** L37–L41 (2003).
31. André, Ph, Ward-Thompson, D. & Barsony, M. From prestellar cores to protostars: the initial conditions of star formation. In *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 59–96 (Univ. Arizona Press, 2000).
32. Klein, B. et al. High-resolution wide-band fast Fourier transform spectrometers. *Astron. Astrophys.* **542,** L3 (2012).
33. Guan, X. et al. GREAT/SOFIA atmospheric calibration. *Astron. Astrophys.* **542,** L4 (2012).
34. Heyminck, S., Kasemann, C., Güsten, R., de Lange, G. & Graf, U. U. The first-light APEX submillimeter heterodyne instrument FLASH. *Astron. Astrophys.* **454,** L21 (2006).
35. Klein, B. et al. The APEX digital Fast Fourier Transform spectrometer. *Astron. Astrophys.* **454,** L29 (2006).
36. Coutens, A. et al. Heavy water stratification in a low-mass protostar. *Astron. Astrophys.* **553,** A75 (2013).
37. Grussie, F. et al. The low-temperature nuclear spin equilibrium of H_3^+ in collisions with H_2. *Astron. Astrophys.* **759,** 21 (2012).
38. Honvault, P., Jorfi, M., González-Lezana, T., Faure, A. & Pagani, L. Ortho-para H_2 conversion by proton exchange at low temperature: an accurate quantum mechanical study. *Phys. Rev. Lett.* **107,** 023201 (2011); Erratum: *Phys. Rev. Lett.* **108,** 109903 (2012).
39. Pagani, L. et al. Chemical modeling of L183/L134N: an estimate of the ortho/para H_2 ratio. *Astron. Astrophys.* **494,** 623–636 (2009).
40. Semenov, D. et al. Chemistry in disks. IV. Benchmarking gas-grain chemical models with surface reactions. *Astron. Astrophys.* **522,** A42 (2010).

41. Albertsson, T. *et al.* First time-dependent study of H_2 and H_3^+ ortho-para chemistry in the diffuse ISM. *Astrophys. J.* **787,** 44 (2014).

42. Crabtree, K. N., Indriolo, N., Kreckel, H., Tom, B. A. & McCall, B. J. On the ortho:para ratio of H_3^+ in diffuse molecular clouds. *Astrophys. J.* **729,** 15 (2011).

43. André, P. *et al.* From filamentary clouds to prestellar cores to the stellar IMF: initial highlights from the Herschel Gould Belt Survey. *Astron. Astrophys.* **518,** L102 (2010).

44. Sadavoy, S. I. *et al.* Class 0 protostars in the Perseus molecular cloud: a correlation between the youngest protostars and the dense gas distribution. *Astrophys. J.* **787,** L18 (2014).
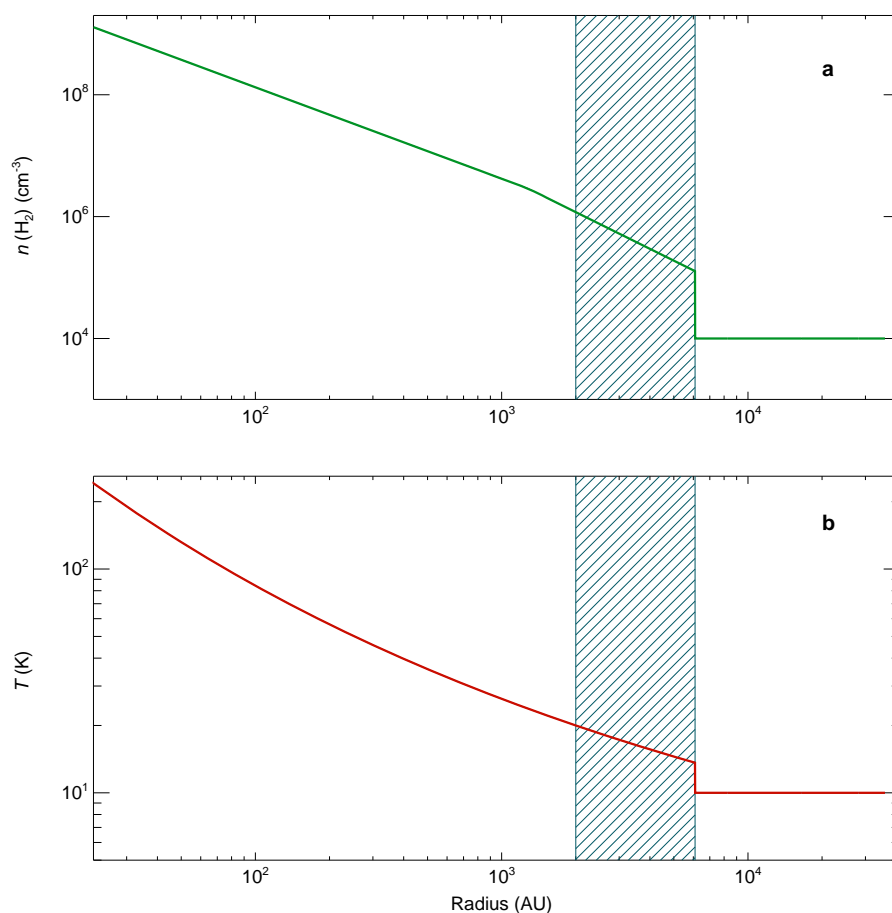
Extended Data Figure 1 | Temperature and density distribution of the source model. Physical model of IRAS 16293-2422 A/B, consisting of a widely used core model[22] and a low-density ambient cloud. **a**, The number density $n(H_2)$ as a function of radius. **b**, The radial profile of the kinetic temperature, T. The ambient cloud is assumed to have $n(H_2) = 10^4$ cm^{-3} and $T = 10$ K. The shaded interval, between a radius of 3,000 and 6,100 AU, represents the outer envelope of the core, which dominates the observed para-H$_2$D$^+$ absorption and ortho-H$_2$D$^+$ emission.
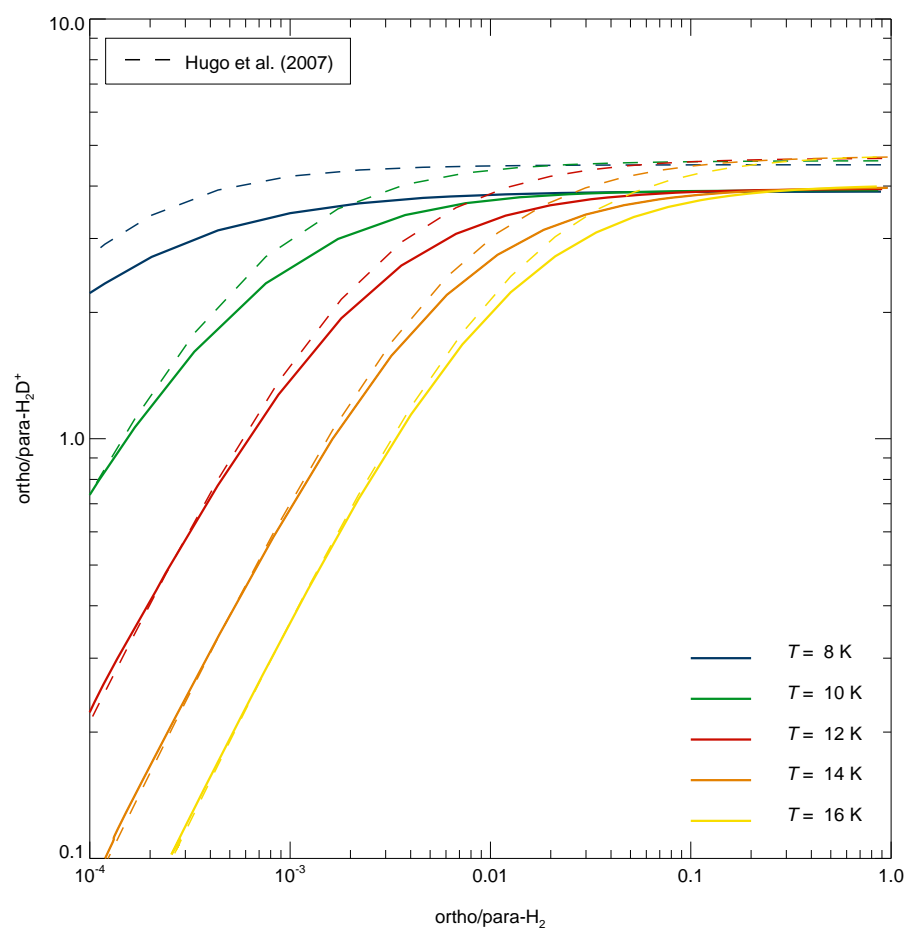
Extended Data Figure 2 | The relationship between ortho/para-H$_2$D$^+$ and ortho/para-H$_2$. The ortho/para-H$_2$D$^+$ ratio as a function of ortho/para-H$_2$ resulting from chemistry simulations for different values of the kinetic temperature T, indicated with colours. The dashed curves represent the approximation given by the analytical formula from Hugo *et al.*[10].
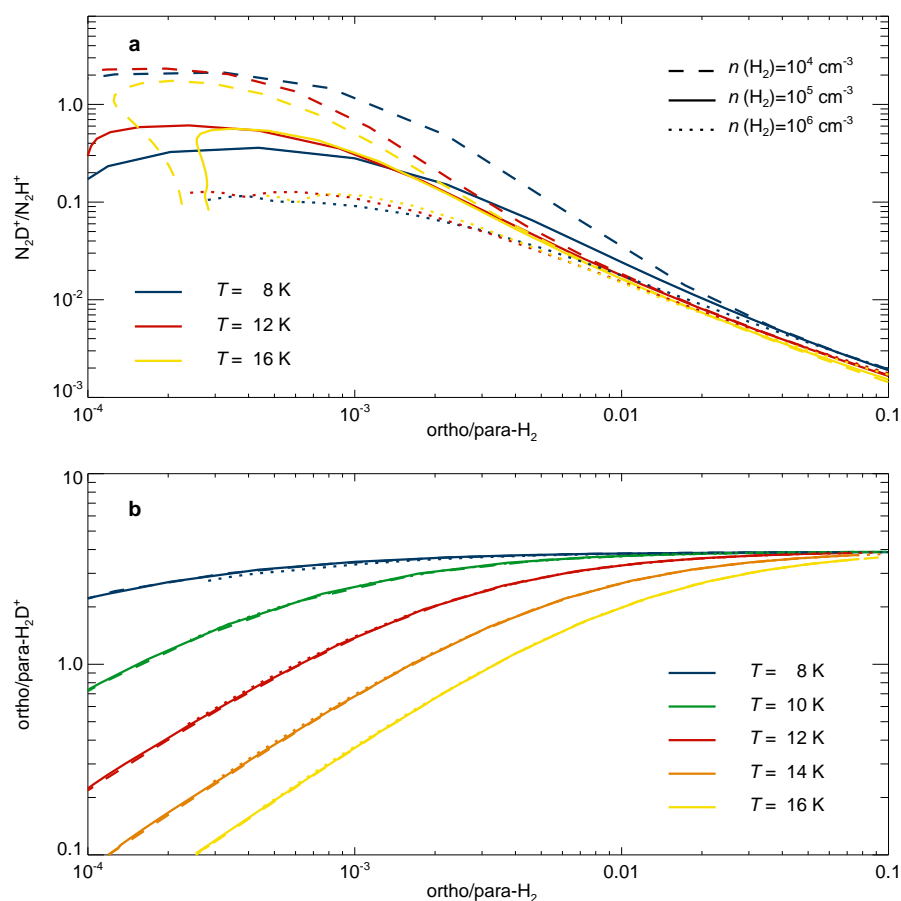
Extended Data Figure 3 | N₂D⁺/N₂H⁺ and ortho/para-H₂D⁺ as functions of ortho/para-H₂, for different values of T and $n(H_2)$. a, The N_2D^+/N_2H^+ abundance ratio versus the ortho/para H_2 ratio for selected values of the kinetic temperature, T, and the H_2 number density, $n(H_2)$. **b**: The ortho/para H_2D^+ ratio versus the ortho/para H_2 ratio for different temperatures and densities. One can see that this relationship depends on T but not on $n(H_2)$.
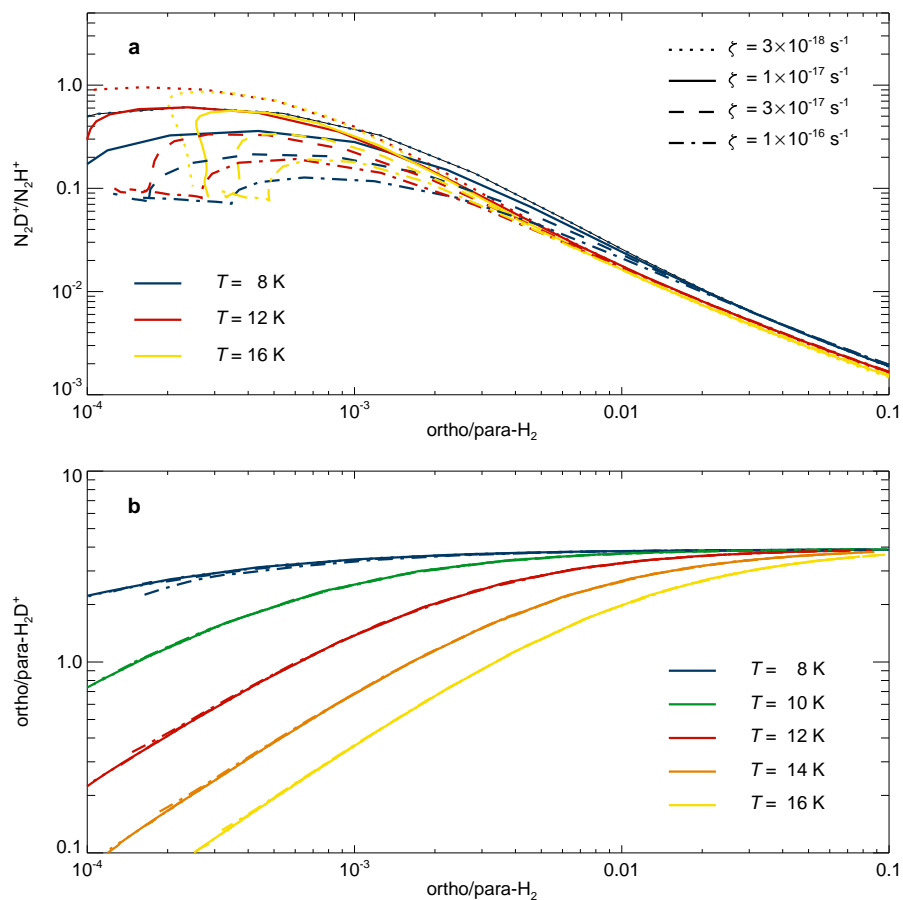
Extended Data Figure 4 | N_2D^+/N_2H^+ and ortho/para-H_2D^+ as functions of ortho/para-H_2, for different values of T and ζ. a, The N_2D^+/N_2H^+ abundance ratio versus the ortho/para H_2 ratio for selected values of the kinetic temperature, T, and the cosmic ray ionization rate, ζ. **b**, The same for the ortho/para H_2D^+ ratio versus the ortho/para H_2 ratio for different temperatures and densities $n(H_2)$. Hardly any dependence on ζ is seen except at the lowest temperatures.
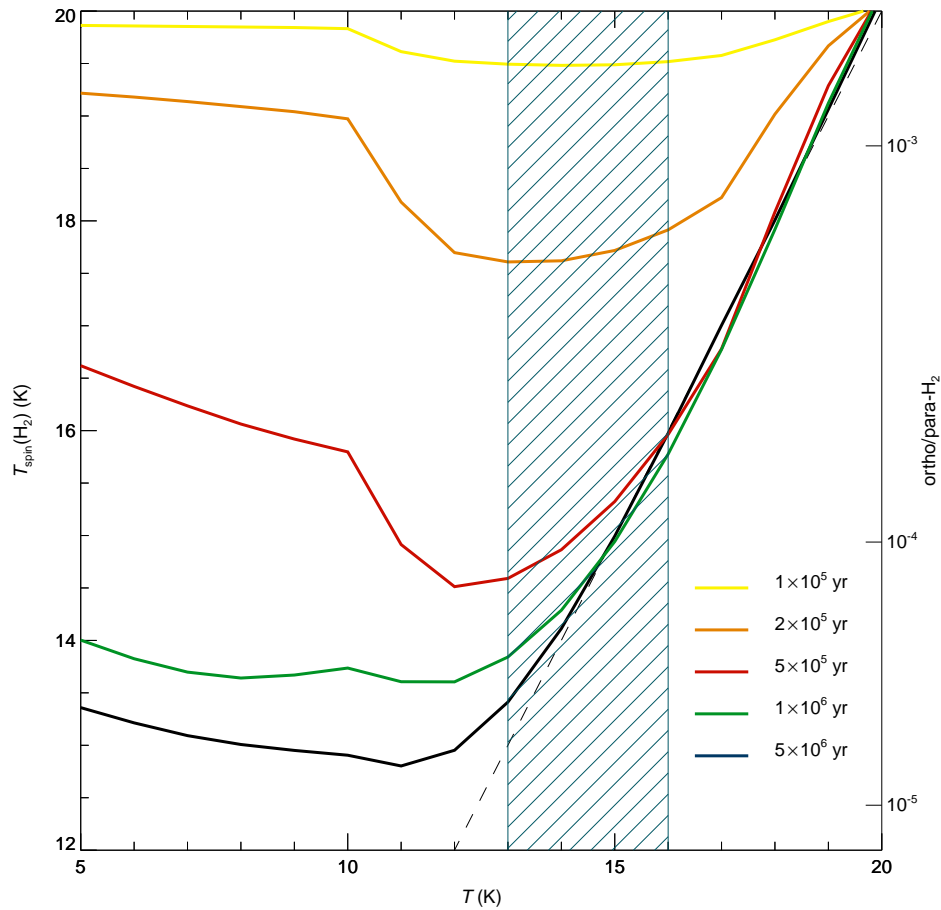
Extended Data Figure 5 | The H₂ spin temperature. Variation of the H$_2$ spin temperature $T_{\rm spin}$ as a function of kinetic temperature and time in a dark cloud according to our gas-grain chemistry model. The corresponding ortho/para-H$_2$ is indicated on the right. The gas density, $n({\rm H}_2) = 10^5\,{\rm cm}^{-3}$, and the visual extinction, $A_{\rm V} = 10$ mag, are kept constant. Ortho/para-H$_2$ tends for long evolutionary times towards the thermal values (dashed line) above $T_{\rm kin} \approx 12$ K. The blue-hatched region indicates the T range applicable to the dense core surrounding IRAS 16293-2422 A/B (between a radius of 3,000 and 6,100 AU).

LETTER

Ultrasensitive mechanical crack–based sensor inspired by the spider sensory system

Daeshik Kang[1,2], Peter V. Pikhitsa[1], Yong Whan Choi[1], Chanseok Lee[1], Sung Soo Shin[1], Linfeng Piao[1], Byeonghak Park[3,4], Kahp-Yang Suh[1,2,5]‡, Tae-il Kim[3,4] & Mansoo Choi[1,2]

Recently developed flexible mechanosensors based on inorganic silicon[1–3], organic semiconductors[4–6], carbon nanotubes[7], graphene platelets[8], pressure-sensitive rubber[9] and self-powered devices[10,11] are highly sensitive and can be applied to human skin. However, the development of a multifunctional sensor satisfying the requirements of ultrahigh mechanosensitivity, flexibility and durability remains a challenge. In nature, spiders sense extremely small variations in mechanical stress using crack-shaped slit organs near their leg joints[12]. Here we demonstrate that sensors based on nanoscale crack junctions and inspired by the geometry of a spider's slit organ can attain ultrahigh sensitivity and serve multiple purposes. The sensors are sensitive to strain (with a gauge factor of over 2,000 in the 0–2 per cent strain range) and vibration (with the ability to detect amplitudes of approximately 10 nanometres). The device is reversible, reproducible, durable and mechanically flexible, and can thus be easily mounted on human skin as an electronic multipixel array. The ultrahigh mechanosensitivity is attributed to the disconnection–reconnection process undergone by the zip-like nanoscale crack junctions under strain or vibration. The proposed theoretical model is consistent with experimental data that we report here. We also demonstrate that sensors based on nanoscale crack junctions are applicable to highly selective speech pattern recognition and the detection of physiological signals. The nanoscale crack junction-based sensory system could be useful in diverse applications requiring ultrahigh displacement sensitivity.

Spiders have crack-shaped slit organs to detect vibrations in their surroundings[12]. The slit geometry enables ultrasensitive displacement detection by allowing for mechanical compliance, which results in the deformation of the slit in response to small external force variations[12,13]. Inspired by this ability, we designed a multifunctional sensor based on nanoscale crack junctions (a 'nanoscale crack sensor') and demonstrated its ultrahigh sensitivity to physiological signals (for example speech patterns and heart rates) and external forces (for example pressure, strain and vibration). The analogy between our nanoscale crack sensor and the spider slit organ is partial because the signal transduction through a spider's neurons and the electrical conduction through our sensor are different. The similarity lies in the slit geometry, which is known to be the key to slit organ ultrasensitivity[12].

An illustration of the spider's slit organ is presented in Fig. 1. The spider has strain detectors located near the leg joint between the metatarsus and tarsus bones[12]. The detectors are composed of a viscoelastic pad, with the slit organ consisting of approximately parallel sensory lyriform embedded in the mechanically stiff exoskeleton (Fig. 1b). The slits are directly connected to the nervous system to collect external vibrations. In this work, we mimicked the geometry of the slit organ to design sensors by depositing a stiff, 20 nm-thick platinum (Pt) layer on top of a viscoelastic polymer, polyurethane acrylate[14] (PUA) (details in 'Experimental section' in Supplementary Information). Analogous to the crack-shaped slit organ, we generated controlled cracks in the Pt film across which electrical conductance

can be measured. The Pt film on PUA was mechanically bent by applying various radii of curvature (1, 2 and 3 mm), and the cracks were formed in a controlled manner in terms of crack density and direction. Studies of controlled crack formation using notches and confined surface stress have been reported[15–17] although cracks were typically considered as a defect to be avoided[15,18]. As shown in Supplementary Fig. 1, the crack spacing (or density) can be controlled by bending the sample with different radii of curvature. The sensor performance is affected by the crack density. The cracked Pt on PUA shown in Fig. 1d has lateral dimensions of 5 mm × 10 mm on 10 μm-thick PUA. Figure 1e illustrates that cracks are formed in the transverse direction to the extension force applied with a bending curvature radius of 1 mm. Supplementary Fig. 2 shows that the cracks penetrate the Pt film and extend into the PUA substrate with a total crack depth of approximately 40–50 nm (ref. 19). The crack gap increases with strain, as shown in Supplementary Fig. 2 and Fig. 1f. Even at 0% strain, a small gap (~5 nm) exists between matching crack edges, indicating that not all of these edges are in contact with each other. A simplified sketch of our nanoscale crack sensor is shown in Fig. 1c. Figure 1g illustrates the widening of the 50 nm-deep crack gap by stretching using finite-element method simulations.

The electrical conductance of a metal strip with a straight transverse cut experiences a sudden jump from a finite value when the edges of the cut are in contact, to zero when they disconnect. For cracks in the Pt film, the high strain sensitivity originates from the rare yet large gap-bridging steps on opposite edges of a zigzag crack. Large variations in resistance are obtained with high repeatability for a cracked sample with a bending curvature radius of 1 mm when the sensor is loaded to produce up to 2% strain and unloaded back to 0% strain at a sweeping speed of 1 mm min^{-1} (Fig. 2a). Figure 2b shows such cyclic variations in resistance for different peak strains, in sharp contrast to the case with a nearly flat bare Pt film with no cracks (yellow curve). The current–voltage ($I–V$) curves for the crack sample and the bare film without cracks are presented in Supplementary Fig. 3 for various strains. The same cyclic measurements performed at a slower sweeping speed of 0.1 mm min^{-1} in Fig. 2c illustrate that the loading and unloading are nearly reversible. When compared with the case with no crack (Fig. 2c, inset), the crack sample exhibits a 450-fold-higher resistance variation (ΔR) at 0.5% strain. We obtained reproducible results from thirty different samples (Supplementary Fig. 4). The durability was confirmed by performing 5,000 cyclic strain tests (Supplementary Fig. 5). As noted earlier, controlled crack formation using different bending curvature radii resulted in different crack spacings (Supplementary Fig. 1), which affected the sensor performance in a controllable manner (Supplementary Fig. 6). The gauge factor determined from the definition[5] ($\Delta R/R_0\varepsilon$) exceeds 2,000 at strains of 0–2% (Supplementary Fig. 7). The strain-dependent gauge factors determined by measuring the derivative of R/R_0 in Fig. 2c were compared with those obtained from the approach to sensor construction based on graphene platelets[8] (Supplementary Fig. 8).

[1]Global Frontier Center for Multiscale Energy Systems, Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-742, South Korea. [2]Division of WCU Multiscale Mechanical Design, Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-742, South Korea. [3]Center for Neuroscience Imaging Research (CNIR), Institute for Basic Science (IBS), Suwon 440-746, South Korea. [4]School of Chemical Engineering, Sungkyunkwan University (SKKU), Suwon 440-746, South Korea. [5]Interdisciplinary Program of Bioengineering, Seoul National University, Seoul 151-742, South Korea.
‡Deceased.
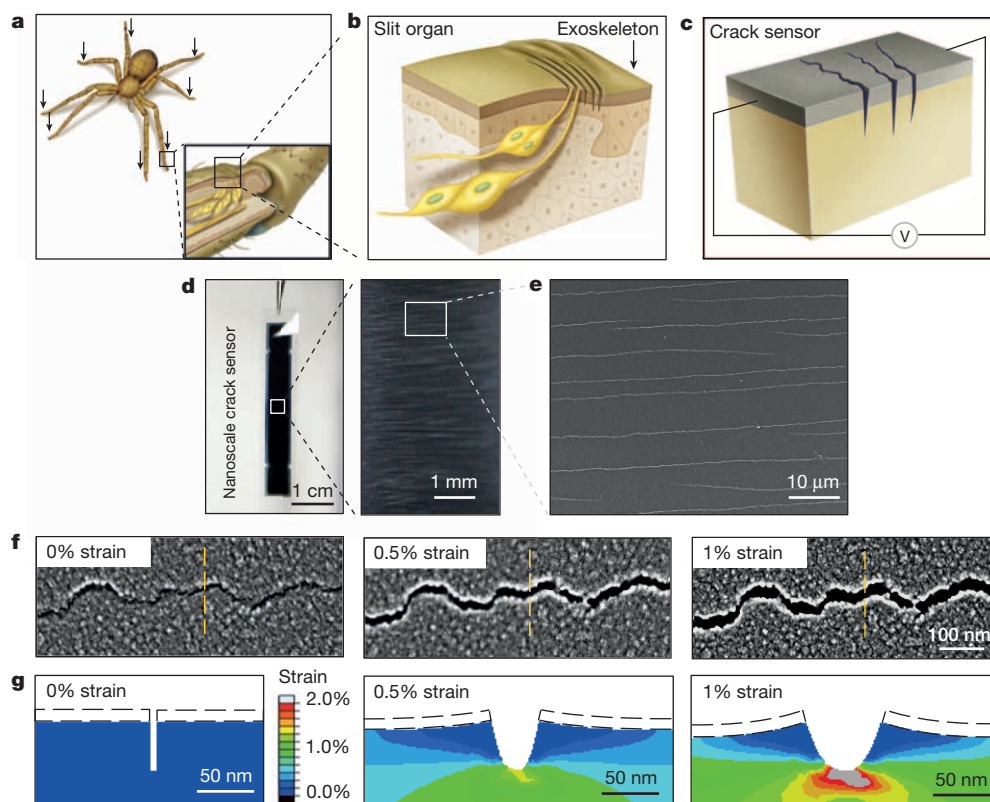
Figure 1 | Schematic illustrations and images of an ultra-mechanosensitive nanoscale crack junction-based sensor inspired by the spider sensory system. a, The spider has highly sensitive organs located on its leg joints (black arrows) for the detection of external forces and vibrations. Inset, enlarged images of the sensory slit organs in the vicinity of the leg joint between the metatarsus and the tarsus. **b**, The slits are connected to the nervous system to monitor vibrations. The slits are in the highly stiff exoskeleton (surface) and a viscoelastic pad (below the exoskeleton). **c**, Illustration of the crack-based sensor and its measurement scheme. Grey, platinum layer; beige, viscoelastic polymer layer. **d**, Left, image of the spider-inspired sensor with a cracked, 20 nm-thick Pt layer formed by bending with a 1 mm radius of curvature. The sensor has lateral dimensions of 5 mm × 10 mm on 10 μm-thick PUA. Right, enlarged image of the cracks in the surface of the sensor in the left-hand image. **e**, SEM image of the boxed region of the right-hand image in **d**. **f**, SEM images of the zip-like crack junctions for different applied strains: 0% (left), 0.5% (middle) and 1% (right). **g**, Finite-element method modelling results of crack interfacial deformation by 0% (left), 0.5% (middle) and 1% (right) strain. The white regions surrounded by black dashes represent the 20-nm-thick Pt layer.

To demonstrate the device's scalability and ability to detect mechanical vibrations and pressure, we used a sensing network of 64 pixels (8 × 8 pixel array) with dimensions of 5 cm × 5 cm (Fig. 2d; details in 'Experimental section' in Supplementary Information and Supplementary Figs 9 and 10). The flexible format of a multipixel array (Fig. 2g) enables the simultaneous measurement of two different stimuli, pressure and vibration, using a simple analyser scheme (Fig. 2h). The results for static pressure applied using a piece of PDMS piece (5 Pa; Fig. 2e) and dynamic pressure simulating a flapping ladybird (5 Pa of pressure and a vibration of frequency 200 Hz and amplitude 14 μm; Supplementary Fig. 11) are shown in Fig. 2k and Fig. 2l, respectively. A piece of PDMS was placed on the red-boxed region in Fig. 2d (see also Fig. 2e) as a static pressure input, and a piezoelectric vibrator was placed on the blue-boxed region in Fig. 2d (see also Fig. 2f) as a vibration source simulating a ladybird's flapping. The distributions of applied pressure from both stimuli can be detected at both locations (Fig. 2i). However, the vibration signal is selectively detected only at the spot where the vibration input is applied (Fig. 2j). Figure 2k, l illustrates the dramatic changes in the *in situ* signals of these pixels at both locations. The applied 200 Hz vibration was obtained by Fourier transform (Fig. 2l, inset). The flexibility of our sensor was examined by measuring the same vibration signals using bent nanoscale crack sensors with different curvature radii (Supplementary Figs 12 and 13).

The nanoscale crack sensor is able to monitor minute vibrations caused by sound waves. To demonstrate its performance as a sound monitor, the sensor was attached to the surface of a violin (Fig. 3a). The sensor measures the strings' vibrations on the right side of the instrument above the f-hole, which allows the resonating air inside the violin to emerge. The measured G-, D-, A- and E-string sounds reveal peak signals at different frequencies that correspond to the known frequencies (Fig. 3b). Time-dependent resistance variations were also measured while Elgar's 'Salut d'Amour' was played, and they were converted into digital signals (Supplementary Video 1). From those signals, the real-time peak spectrogram was retrieved (Fig. 3c). The harmonic frequency of each note is recorded correctly.

A flexible sensor attached to a human neck can be used as a speech pattern recognition system. A microphone-based system cannot filter unnecessary information in a noisy environment, in contrast to the human auditory system (known as the 'cocktail party phenomenon')[20]. We asked ten human speakers to repeat four simple words ('go', 'jump', 'shoot' and 'stop') more than ten times with the crack sensor attached to their necks (Fig. 3f) and in front of a standing microphone (Supplementary Fig. 14). The acoustic waveforms and auditory spectrograms of the human speakers were analysed by real-time fast Fourier transform. In silence, the acoustic waveforms (Fig. 3d, top) and their respective spectrograms (Fig. 3d, bottom) from both tools, the nanoscale crack sensor (blue) and the standing microphone (red), are stable. However, in a noisy environment of approximately 92 dB (measured using a Brüel & Kjaer Type 2250 sound level meter), the spectrogram from the nanoscale crack sensor (Fig. 3e, green) remains stable, whereas that from the standing microphone (Fig. 3e, black) becomes noisy. We also tested the commercially available CMP-756 electret condenser microphone (CUI Inc.) while it was attached to a speaker's neck to compare the accuracy of word recognition (Supplementary Fig. 14). The accuracy of simple word recognition for our nanoscale crack sensor was approximately
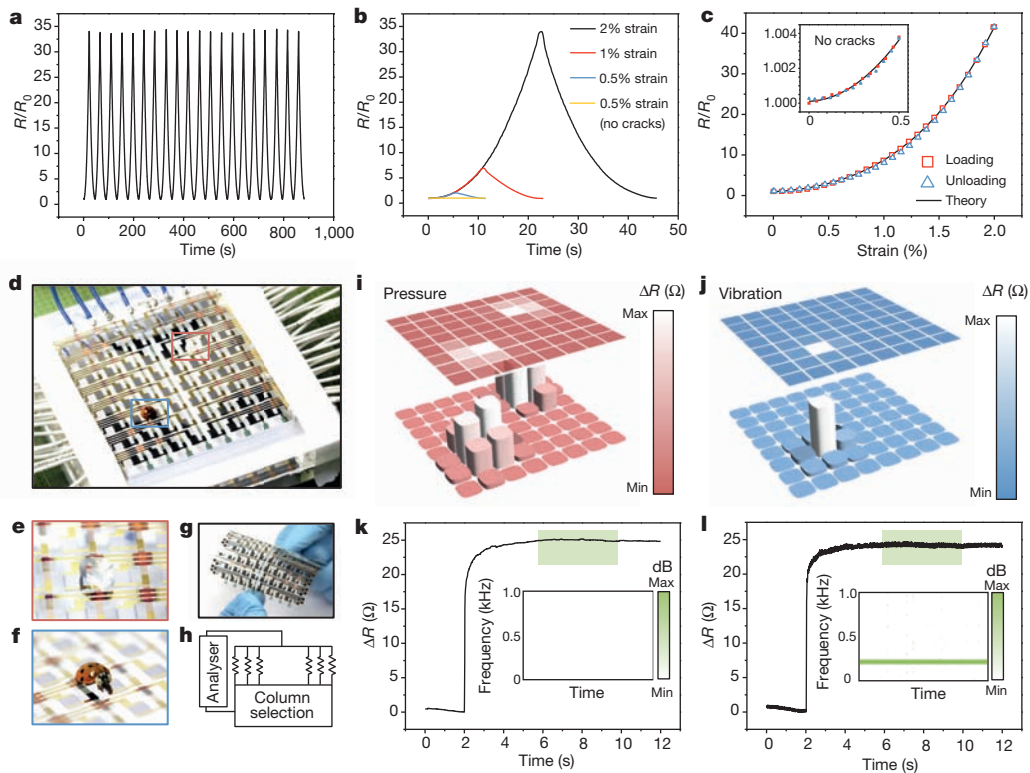
Figure 2 | Resistance variations with strain and the multipixel array of the crack sensor. a, The normalized resistance measured at a strain sweep rate of 1 mm min^{-1}. **b,** Reversible loading–unloading behaviour for various final strains. **c,** Resistance at the slowest loading–unloading rate, of 0.1 mm min^{-1}, compared with the theoretical fit. Inset, results for no cracks. **d,** The 8×8 array of the crack sensor. Pressure was applied with a piece of polydimethylsiloxane (PDMS; red), and vibration and pressure were applied using a flapping ladybird (blue). The overall dimensions of the device are 5 cm $\times$ 5 cm, and each pixel is 2 mm $\times$ 2 mm. **e,** Region where pressure was applied using PDMS.

f, Region where pressure and 0.2 kHz vibration were applied using a flapping ladybird. **g,** Representative image of the nanoscale crack sensor's flexibility. **h,** Simple circuit scheme of the array in **d** for multiplexing. **i,** Pressure distribution with a piece of PDMS and a non-flapping ladybird. **j,** Vibration distribution with a piece of PDMS and a flapping ladybird. **k,** Dynamic pressure change in the red box in **d**. Inset, no vibration measured in the green shaded region. **l,** Dynamic pressure change with 0.2 kHz vibrations. Inset, frequency of 0.2 kHz measured in the green shaded region.

97.5% even with noise. Another test was done to confirm that our sensor could successfully pick up complicated voice patterns from a song by attaching our sensor to the diaphragm of a loudspeaker while the song was played in a noisy environment.

Figure 3g, h presents another example in which we measured heart rates under two different conditions, normal and after running. The signals were successfully monitored *in situ* and provide crucial heart physiology information, such as the diastolic and systolic movements of the heart (Supplementary Fig. 15). To demonstrate another application, the nanoscale crack sensor was integrated into a microfluidic system to measure the input flow rate by showing the linear variation of resistance change with flow rate (Fig. 3i and Supplementary Fig. 16). The results of sensing a sinusoidal/step-function force and a 5 μl water droplet are shown in Supplementary Fig. 17.

To investigate the sensor mechanism, we studied the normalized conductance, $S = R_0/R$, as a function of strain (Supplementary Fig. 18). This revealed an intriguing fluctuating behaviour, particularly at lower strains (Supplementary Fig. 18, inset). The derivative $-dS/d\varepsilon$ displays large fluctuations with negative and positive values, particularly at strains of less than 1% (Fig. 4a). These fluctuations are well beyond the noise level observed for the bare film without cracks (Fig. 4a, inset). We attribute these intriguing fluctuations to the disconnection–reconnection events of the crack edges. A positive $-dS/d\varepsilon$ value represents a disconnection event whereas a negative $-dS/d\varepsilon$ value represents a reconnection. A cracked film over an elastic substrate with a positive Poisson's ratio could be compressed in the transverse direction while being extended in the axial direction. This indicates that the axial extension could disconnect the crack edges and that the lateral compression could reconnect them. In Fig. 4a, there are two distinct strain

regions with the larger strain region being characterized by only positive fluctuations. This confirms that the larger steps in the crack edges preferably disconnect under loading. At lower strains, the fluctuations are both positive and negative, indicating disconnections and reconnections for numerous small steps in the crack edges. Averaging the positive and negative spikes ($\langle -dS/d\varepsilon \rangle$) (Fig. 4b, red and grey curves) yields a positive value in all areas, indicating that the net effect of disconnection–reconnection is to reduce conductance as the extension proceeds. A further detailed description of the disconnection–reconnection process is provided in Supplementary Fig. 19. This overall behaviour of $\langle -dS/d\varepsilon \rangle$ is related to the crack asperity size distribution because the disconnection–reconnection events should depend on the crack asperity size distribution. Dynamic sweeping motion results in sweeping rate-dependent resistance variations, although the curves are nearly reversible (Supplementary Fig. 20). The sweeping rate-dependent resistance variation is attributed to the rate-dependent nature of the crack disconnection–reconnection process (Supplementary Fig. 20).

For uniaxial strain, the elastic strip becomes compressed transversally, and small edge steps remain in contact until the strain completely disconnects them. This process occurs when the gap distance overcomes the crack asperity height (in the simplified diagram in Fig. 4c, d, the height of two blue grains is defined as the crack asperity height, with each grain representing a small step). Scanning electron microscope (SEM) images illustrate that the gap distance is proportional to the strain: $d = k\varepsilon$, where $k \approx 70$ nm and ε is in per cent (Supplementary Fig. 21). A central component of the mechanism of conduction across a crack is a simplified expression for S that accounts for the sudden termination of a contact when the gap $k\varepsilon$ exceeds $h_i = k\varepsilon_i$, the height of the crack's ith asperity
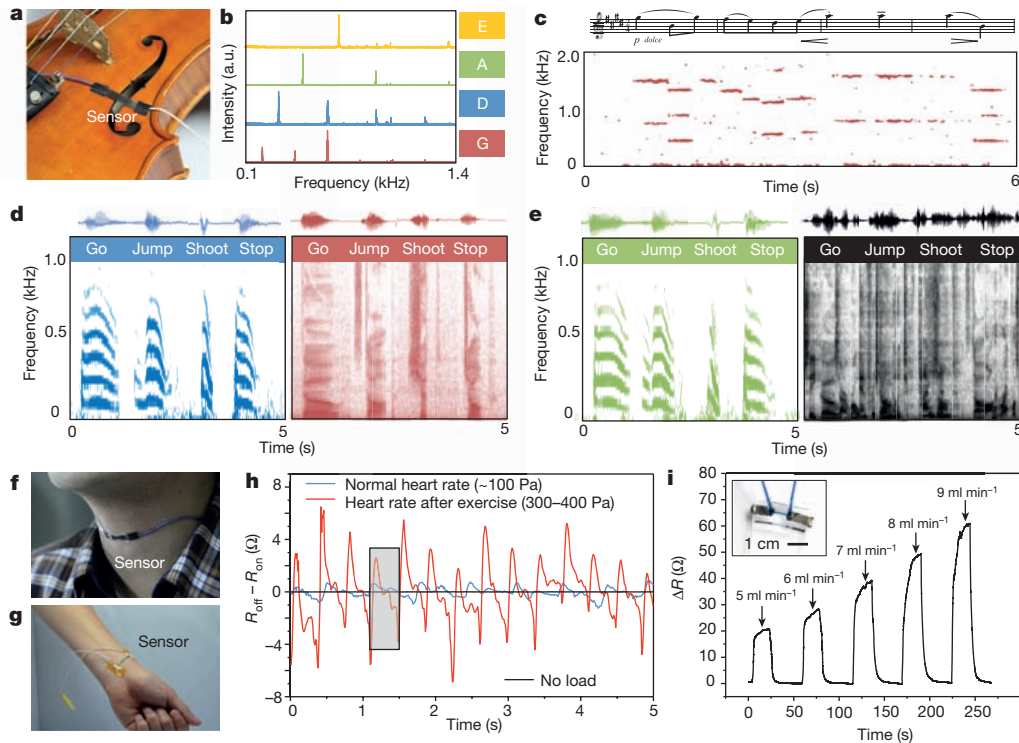
Figure 3 | Nanoscale crack junction-based sensor applications for sound and speech pattern recognition, human physiology monitoring and flow rate indicators. a, Image of a nanoscale crack sensor attached to a violin for sound wave recognition. The device is placed on the right string above the f-hole of the violin using commercial tape. **b**, E (yellow), A (green), D (blue) and G (red) strings played open (that is, with no finger stopping) produce different wavefunctions, which we collected using the nanoscale crack sensor. The E, A, D and G strings have fundamental frequencies of 659, 440, 294 and 196 Hz, respectively, as measured by the sensor. **c**, The measured sound waves of music playing (Salut d'Amour; excerpt shown at top). **d**, **e**, Comparisons of the acoustic waveform and auditory spectrogram changes measured by electrical resistance using the nanoscale crack sensor (left-hand images) and a standing microphone (right-hand images) in quiet (**d**) and noisy (**e**; ~92 dB) environments. All of the signals are measurements of a person saying 'go', 'jump', 'shoot' and 'stop'. The signals from both the nanoscale crack sensor and

the microphone are recorded clearly without noise (**d**). The signal from the standing microphone is not clear with ~92 dB of noise (right-hand image in **e**), whereas our crack sensor maintains its high level of accuracy under the same noise level (left-hand image in **e**). **f**, Image of the nanoscale crack sensor attached to a person's neck for human speech recording. **g**, Image of the nanoscale crack sensor attached to a person's wrist for pulse measurement. **h**, Measured characteristics of the resistance difference for the nanoscale crack sensor attached to a person's wrist. The detailed variations of the pulses for the reference (no load; black), normal heart rate (load of ~100 Pa; blue) and heart rate after running (300–400 Pa; red) are clearly observed. **i**, Resistance change at various flow rates as a function of time, measured using the nanoscale crack sensor encapsulated by a PDMS spacer in a microchannel. Inset, image of the nanoscale crack sensor attached to a microfluidic channel for liquid flow rate measurement.

peak (Supplementary Fig. 21). The crack surfaces do not all touch the opposite side; judging from the magnified SEM images in Supplementary Fig. 21b, only a small number of contacts exist. However, considering the width (~5 mm) and density (~1,000 cm^{-1}) of the cracks, many (of order 10^5) opposing crack surfaces are in contact in each sensor, and these crack surfaces lead to variations in the conductance.

Considering the above process, the simplified form of the normalized conductance can be written as

$$S = \frac{\sum_i N_i \theta(\varepsilon_i - \varepsilon)}{\sum_i N_i} \quad (1)$$

where $\theta(\varepsilon_i - \varepsilon)$ is the Heaviside step function and N_i is the number of crack asperities of height $k\varepsilon_i$. For a normalized probability distribution function of crack asperity size $p(\varepsilon)$, we rewrite equation (1) as

$$S = \int_\varepsilon^\infty p(y) \, \mathrm{d}y. \quad (2)$$

We argue that the small variations in crack asperity due to grain shifts are distributed in the same manner as the large variations due to grain piling, which yields an equation for $p(\varepsilon)$ as a log-normal distribution function

(details in 'Theory section' in Supplementary Information; ε_0 and μ are fitting parameters):

$$p(\varepsilon) = \frac{\exp\left(-\ln(\varepsilon/\varepsilon_0)^2/\mu^2\right)}{\varepsilon\mu\sqrt{\pi}} \quad (3)$$

Crack asperity heights have previously been approximated by a log-normal distribution[21]. Combining equations (3) and (2) yields

$$S = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{\ln(\varepsilon/\varepsilon_0)}{\mu}\right)\right) \quad (4)$$

where $\mathrm{erf}(x)$ is the error function. Equation (4) provides the resistance, $R = 1/S$, which well fits the experimental data shown in Fig. 2c. The experimental values for $-\mathrm{d}S/\mathrm{d}\varepsilon$ averaged over different numbers of data points agree well with the theoretical $-\mathrm{d}S/\mathrm{d}\varepsilon$ obtained from equation (4) (Fig. 4b). The size distribution of the crack asperity heights ($p(\varepsilon)$) was measured from 50 SEM images and is presented in Fig. 4b for comparison with the log-normal distribution (equation (3)) and the experimental average $\langle -\mathrm{d}S/\mathrm{d}\varepsilon\rangle$ because $p(\varepsilon)$ should be equal to $-\mathrm{d}S/\mathrm{d}\varepsilon$ according to our theoretical model. The crack asperity heights also have a long-tailed skewed distribution that is consistent with equation (3) and $\langle -\mathrm{d}S/\mathrm{d}\varepsilon\rangle$ for large strains. The large discrepancy at small strains is attributed to the fact that an initial gap of 5–10 nm exists even at 0% strain (Fig. 1f and Supplementary Fig. 21); thus, the presence of many small crack asperities with magnitudes less than the initial gap does not cause variations in the electrical
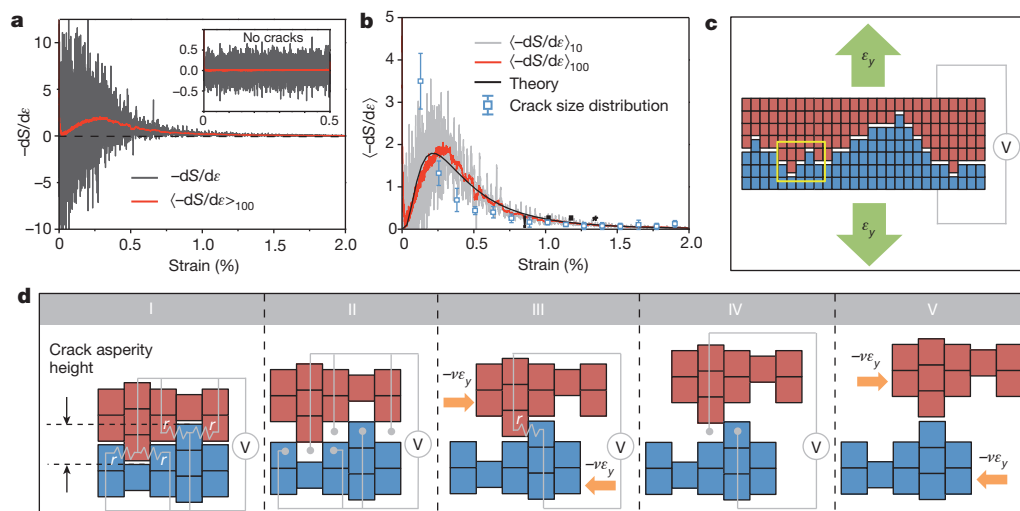
Figure 4 | Theoretical analysis of the nanoscale crack sensor. a, Negative of the derivative of the conductance (grey) and its average (red). Inset, results for no cracks. **b**, Average of the negative of the conductance derivative, and comparison between the theoretical fit from equation (3) ($\varepsilon_0 = 0.4$ and $\mu = 0.98$) and the average derivative from **a**. $\langle -dS/d\varepsilon \rangle_{10}$ and $\langle -dS/d\varepsilon \rangle_{100}$ are the averages for 10 and 100 data points, respectively. The maximum at approximately 0.3% strain corresponds to $0.3k = 21$ nm, which is close to the

constituent primary grain size. The error bars show the size distribution of the crack asperity heights derived from SEM images of the crack edges. **c**, Illustration of the matching crack edges. **d**, Diagram of the behaviour of the boxed, square-grain crack asperity in **c** during the stages of disconnection–reconnection along the step under vertical loading strain and horizontal compression (orange arrows) with a Poisson's ratio of $v \approx 0.3$. r, reconnection.

conductance. A different Pt film thickness was also studied to illustrate that the hysteresis loops are clearly pronounced for a 100 nm-thick Pt film with cracks (see Supplementary Fig. 22 for an explanation of the hysteresis of thick films). A 20 nm-thick Au film was also studied, and the same bending with a 1 mm curvature radius was performed. Unlike the Pt film, the Au film did not generate similar straight cracks. Both the as-prepared Au film and the bent film exhibited random island-type cracks (Supplementary Fig. 23) and maintained conductivity while the film was stretched[22–24] (see Supplementary Table 1 for a comparison of the gauge factors). This indicates that nearly cut-through straight cracks with nanoscale, jagged crack edges, similar to those in our Pt film, would be required to provide the demonstrated ultrasensitivity.

Precise engineering of controlled crack formation other than the method of bending that we used here may further improve the performance of our crack-based ultrasensitive mechanosensor.

1. Kim, D.-H. et al. Epidermal electronics. Science **333**, 838–843 (2011).
2. Takei, K. et al. Nanowire active-matrix circuitry for low-voltage macroscale artificial skin. Nature Mater. **9**, 821–826 (2010).
3. Kim, D.-H., Lu, N., Huang, Y. & Rogers, J. A. Materials for stretchable electronics in bioinspired and biointegrated devices. MRS Bull. **37**, 226–235 (2012).
4. Kaltenbrunner, M. et al. An ultra-lightweight design for imperceptible plastic electronics. Nature **499**, 458–463 (2013).
5. Chortos, A. et al. Highly stretchable transistors using a microcracked organic semiconductor. Adv. Mater. **26**, 4253–4259 (2014).
6. Someya, T. et al. A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications. Proc. Natl Acad. Sci. USA **101**, 9966–9970 (2004).
7. Yamada, T. et al. A stretchable carbon nanotube strain sensor for human-motion detection. Nature Nanotechnol. **6**, 296–301 (2011).
8. Li, X. et al. Stretchable and highly sensitive graphene-on-polymer strain sensors. Sci. Rep. **2**, 870 (2012).
9. Mannsfeld, S. C. et al. Highly sensitive flexible pressure sensors with microstructured rubber dielectric layers. Nature Mater. **9**, 859–864 (2010).
10. Wang, Z. L. Self-powered nanosensors and nanosystems. Adv. Mater. **24**, 280–285 (2012).
11. Wu, W., Wen, X. & Wang, Z. L. Taxel-addressable matrix of vertical-nanowire piezotronic transistors for active and adaptive tactile imaging. Science **340**, 952–957 (2013).
12. Fratzl, P. & Barth, F. G. Biomaterial systems for mechanosensing and actuation. Nature **462**, 442–448 (2009).
13. Hößl, B., Böhm, H., Rammerstorfer, F., Müllan, R. & Barth, F. Studying the deformation of arachnid slit sensilla by a fracture mechanical approach. J. Biomech. **39**, 1761–1768 (2006).
14. Choi, S.-J., Yoo, P. J., Baek, S. J., Kim, T. W. & Lee, H. H. An ultraviolet-curable mold for sub-100-nm lithography. J. Am. Chem. Soc. **126**, 7744–7745 (2004).
15. Nam, K. H., Park, I. H. & Ko, S. H. Patterning by controlled cracking. Nature **485**, 221–224 (2012).
16. Kim, B. C. et al. Guided fracture of films on soft substrates to create micro/nano-feature arrays with controlled periodicity. Sci. Rep. **3**, 3027 (2013).
17. Vandeparre, H., Liu, Q., Minev, I. R., Suo, Z. & Lacour, S. P. Localization of folds and cracks in thin metal films coated on flexible elastomer foams. Adv. Mater. **25**, 3117–3121 (2013).
18. Livne, A., Bouchbinder, E., Svetlizky, I. & Fineberg, J. The near-tip fields of fast cracks. Science **327**, 1359–1363 (2010).
19. Thouless, M., Li, Z., Douville, N. & Takayama, S. Periodic cracking of films supported on compliant substrates. J. Mech. Phys. Solids **59**, 1927–1937 (2011).
20. Bronkhorst, A. W. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acta Acust. United Acust. **86**, 117–128 (2000).
21. Drury, W., Gokhale, A. M. & Antolovich, S. Effect of crack surface geometry on fatigue crack closure. Metall. Mater. Trans. A **26**, 2651–2663 (1995).
22. Lacour, S. P., Wagner, S., Huang, Z. & Suo, Z. Stretchable gold conductors on elastomeric substrates. Appl. Phys. Lett. **82**, 2404–2406 (2003).
23. Lacour, S. P., Jones, J., Wagner, S., Li, T. & Suo, Z. Stretchable interconnects for elastic electronic surfaces. Proc. IEEE **93**, 1459–1467 (2005).
24. Graz, I. M., Cotton, D. P. & Lacour, S. P. Extended cyclic uniaxial loading of stretchable gold thin-films on elastomeric substrates. Appl. Phys. Lett. **94**, 071902 (2009).

Author Contributions D.K., M.C., T.-i.K. and K.-Y.S. designed the experiments; D.K., Y.W.C., C.L., S.S.S., L.P. and B.P. performed the experiments; K.Y.S., T.-i.K. and M.C. led the work; P.V.P., D.K. and M.C. developed the theory; and D.K., P.V.P., M.C. and T.-i.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C. (mchoi@snu.ac.kr) or T.-i.K. (taeilkim@skku.edu).

LETTER

Proton transport through one–atom–thick crystals

S. Hu[1,2], M. Lozada-Hidalgo[1], F. C. Wang[3], A. Mishchenko[1], F. Schedin[2], R. R. Nair[1], E. W. Hill[2], D. W. Boukhvalov[4], M. I. Katsnelson[4], R. A. W. Dryfe[5], I. V. Grigorieva[1], H. A. Wu[3] & A. K. Geim[1,2]

Graphene is increasingly explored as a possible platform for developing novel separation technologies[1–19]. This interest has arisen because it is a maximally thin membrane that, once perforated with atomic accuracy, may allow ultrafast and highly selective sieving of gases, liquids, dissolved ions and other species of interest[2,9–19]. However, a perfect graphene monolayer is impermeable to all atoms and molecules under ambient conditions[1–7]: even hydrogen, the smallest of atoms, is expected to take billions of years to penetrate graphene's dense electronic cloud[3–6]. Only accelerated atoms possess the kinetic energy required to do this[20,21]. The same behaviour might reasonably be expected in the case of other atomically thin crystals[22,23]. Here we report transport and mass spectroscopy measurements which establish that monolayers of graphene and hexagonal boron nitride (hBN) are highly permeable to thermal protons under ambient conditions, whereas no proton transport is detected for thicker crystals such as monolayer molybdenum disulphide, bilayer graphene or multilayer hBN. Protons present an intermediate case between electrons (which can tunnel easily through atomically thin barriers[24]) and atoms, yet our measured transport rates are unexpectedly high[4,5] and raise fundamental questions about the details of the transport process. We see the highest room-temperature proton conductivity with monolayer hBN, for which we measure a resistivity to proton flow of about $10\,\Omega\,cm^2$ and a low activation energy of about 0.3 electronvolts. At higher temperatures, hBN is outperformed by graphene, the resistivity of which is estimated to fall below $10^{-3}\,\Omega\,cm^2$ above 250 degrees Celsius. Proton transport can be further enhanced by decorating the graphene and hBN membranes with catalytic metal nanoparticles. The high, selective proton conductivity and stability make one-atom-thick crystals promising candidates for use in many hydrogen-based technologies.

We have investigated the possibility of proton transport through monocrystalline membranes made from mono- and few-layer graphene, hBN, and molybdenum disulphide (MoS_2). The two-dimensional (2D) crystals[22,23] were obtained by micromechanical cleavage and then suspended over micrometre-size holes etched through Si wafers (Extended Data Figs 1 and 2). The resulting free-standing membranes were checked for the absence of pinholes and defects and were coated on both sides with Nafion, a polymer with high proton conductivity and negligible electron conductivity[25]. Finally, two proton-injecting PdH_x electrodes[26,27] were deposited onto the Nafion from both sides of the wafer. (See 'Experimental devices' in Methods for a detailed description of the fabrication procedures.) As illustrated in the left inset of Fig. 1a, the 2D crystals effectively serve as atomically thin barriers between two Nafion spaces. For electrical measurements ('Conductance measurements' in Methods), samples were placed in a H–Ar atmosphere at 100% humidity, which ensured high conductivity of the Nafion films[25,26]. Examples of current–voltage characteristics measured for devices incorporating monolayers of graphene, hBN and MoS_2 are shown in Fig. 1a. The measured proton current I varies linearly with bias voltage V, with conductance $S = I/V$ proportional to the membrane area A (Extended Data
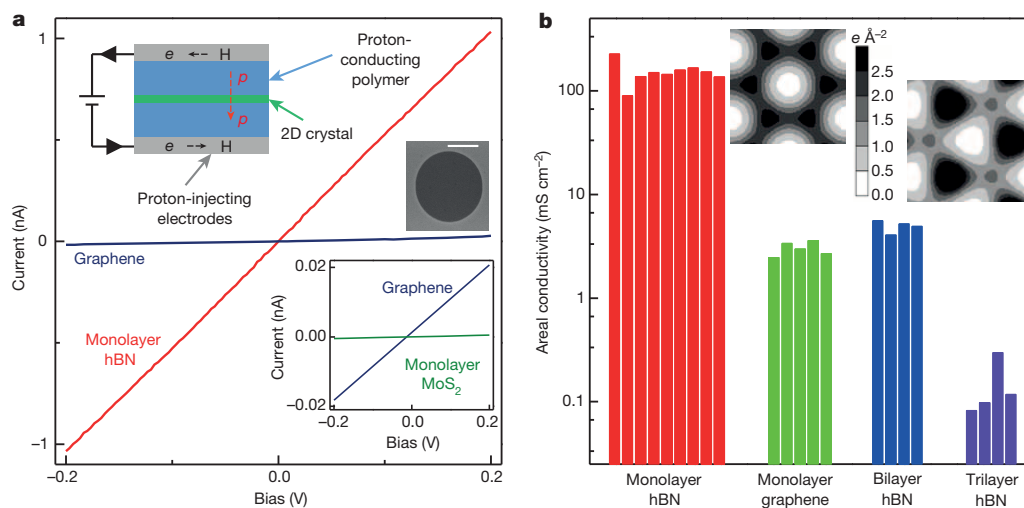
Figure 1 | Proton transport through 2D crystals. a, Examples of I–V characteristics for monolayers of hBN, graphite and MoS_2. The upper inset shows a sketch of the experimental set-up. The middle inset (scale bar, 1 μm) shows an electron micrograph of a typical graphene membrane before the deposition of Nafion. Small (pA) currents observed for MoS_2 membrane devices (lower inset) are due to parasitic parallel conductance. **b**, Histograms for 2D crystals that are found to exhibit measurable proton conductivity. Each bar represents a different sample with a 2 μm-diameter membrane. Insets, charge density (in electrons per $Å^2$) integrated along the direction perpendicular to graphene (left) and monolayer hBN (right). The white areas are minima at the hexagon centres; the maxima correspond to positions of C, B and N atoms. The measurements were carried out at room temperature (21–23 °C).

[1]School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK. [2]Manchester Centre for Mesoscience and Nanotechnology, The University of Manchester, Manchester M13 9PL, UK. [3]Chinese Academy of Sciences Key Laboratory of Mechanical Behavior and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei, Anhui 230027, China. [4]Institute for Molecules and Materials, Radboud University of Nijmegen, 6525 AJ Nijmegen, The Netherlands. [5]School of Chemistry, The University of Manchester, Manchester M13 9PL, UK.

Fig. 3). For 'bare-hole' devices, which were prepared in the same manner but lacked a 2D membrane, S was ~50 times higher than in the presence of monolayer hBN. This confirms that the measured areal conductivity $\sigma = S/A$ is dominated by the 2D crystals, with Nafion contributing only a relatively small series resistance. For devices with thick barriers (for example 100 nm-thick metal or insulating films evaporated between the Nafion spaces), we find a parasitic parallel conductance of ~10 pS caused by leakage currents along silicon nitride surfaces at high humidity (Methods). Within this uncertainty, we could not detect any proton current through monolayer MoS_2, bilayer graphene, four-layer hBN or thicker 2D crystals. The reported behaviour was highly reproducible, as illustrated by statistics in Fig. 1b and Extended Data Fig. 4 for a number of different devices. To further demonstrate the generality of the observed behaviour, we also used a set-up where 2D membranes separate liquid electrolyte cells (containing HCl solutions) instead of Nafion (Methods). We found the same proton conductivities using this electrolyte set-up (Extended Data Fig. 5).

Insight into the difference in permeation through different 2D crystals is gained by considering the electron clouds passed by translocating protons, as shown for graphene and monolayer hBN in the insets of Fig. 1b. Monolayer hBN is more 'porous' than graphene, reflecting that the BN bond is strongly polarized, with valence electrons concentrated around N atoms. The non-permeable MoS_2 consists of three atomic layers containing large atoms, resulting in a much denser electron cloud (Extended Data Fig. 6). The absence of detectable σ for bilayer graphene can be attributed to its AB stacking (the hexagonal rings in each graphene layer are centred on the carbon atoms in the adjacent layer). This results in 'pores' in the electron cloud of one layer being covered by electron density maxima within the adjacent layer. In contrast, the AA′ stacking of hBN (hexagonal rings in different layers are aligned with each other) results in an increase in the integrated electron density with increasing layer number but retains the central pore in the electron cloud even for multilayer hBN membranes.

There is no correlation between proton transport and either the electron transport behaviour or the quality of the 2D crystals. hBN exhibits the highest proton conductivity but is a wide-gap insulator with the highest electron tunnelling barrier[23,24], whereas monolayer MoS_2 shows no discernible proton permeation but is a heavily doped semiconductor with electron-type conductivity[22,28]. And whereas extensive examination using transmission and tunnelling electron microscopy and other techniques ('Absence of atomic-scale defects' in Methods) failed to find

even individual pinholes (atomic-scale defects) in graphene and hBN prepared using the same cleavage technique as employed in the present work (see also refs 1, 2, 24 and Extended Data Fig. 7), MoS_2 monolayers contain a high density of sulphur vacancies[29] yet exhibited little proton conductivity. These observations, the high reproducibility of our measurements for different devices, the linear scaling with area A, and the expected changes with increasing layer number all support our conclusion that the measured σ values represent the intrinsic proton conductivity of the studied 2D crystals. (See 'Absence of atomic-scale defects' in Methods for further evidence against the involvement of atomic-scale defects in the observed proton permeation.)

The transport barrier heights E for different 2D crystals are obtained by measuring σ as a function of temperature T (Fig. 2a), revealing that proton conductivities exhibit Arrhenius-type behaviour, $\exp(-E/k_B T)$, where k_B is the Boltzmann constant. We note that the conductivity of Nafion contributes little to the overall value of S, and changes only by a factor of two over the T range examined (Extended Data Fig. 8). The data in Fig. 2a yield $E = 0.78 \pm 0.03$, 0.61 ± 0.04 and 0.3 ± 0.02 eV for graphene, bilayer hBN and monolayer hBN, respectively. Measurements on different devices give values that are reproducible within our experimental accuracy of ~10% (Extended Data Fig. 4). This is consistent with the high reproducibility of σ found for different devices (Fig. 1b) because otherwise different E values should yield hugely different σ values at a given T.

The barrier to proton transport through graphene we have determined is notably lower than the 1.2–2.2 eV found in *ab initio* molecular dynamics simulations and calculations using the climbing-image nudged elastic band method[4–6], which would result in proton conductivities millions of times smaller and undetectable in our experiments. We have reproduced the earlier barrier calculations for graphene and extended them to monolayer hBN ('Theoretical analysis of proton transport through 2D crystals' in Methods), obtaining values of $E = 1.25–1.40$ eV for graphene, in agreement with refs 4, 5, and ~0.7 eV for monolayer hBN. The disagreement between experiment and theory in the absolute value of E is perhaps not surprising given the complex nature of possible transport pathways and the sensitivity of the calculations to pseudopotentials, the exchange correlation functional and so on. The difference might also arise because protons in Nafion and water move along hydrogen bonds[25] rather than in vacuum as assumed by theory so far.

Some applications call for very high proton conductivities, an example being hydrogen fuel cells that require membranes with $\sigma > 1\,S\,cm^{-2}$.
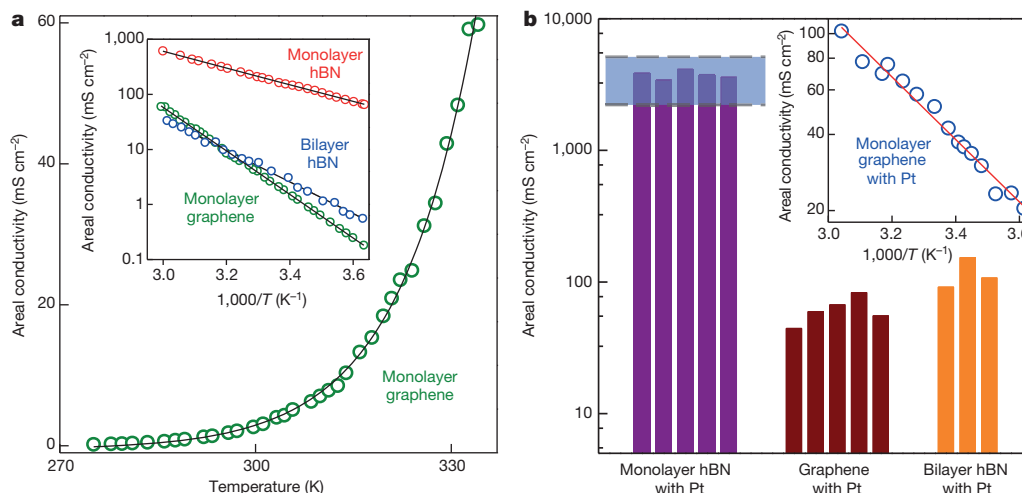
Figure 2 | Proton barrier heights and their catalytic suppression.
a, Temperature dependences of proton conductivity for 2D crystals. The inset shows $\log(\sigma)$ as a function of T^{-1}. Symbols are experimental data; solid curves are the best fits to the activation dependence. The T range is limited by freezing of water in Nafion, and we normally avoided $T > 60\,°C$ to prevent accidental damage because of different thermal expansion coefficients.

b, Proton conductivity of 2D crystals decorated with catalytic nanoparticles. Each bar is a different device. The shaded area shows the conductivity range found for bare-hole devices (Methods). Inset, Arrhenius-type behaviour for graphene decorated with Pt, yielding $E \approx 0.24$ eV. Monolayer hBN decorated with Pt exhibits only a weak T dependence (Extended Data Fig. 8), which indicates that its E becomes comparable to $k_B T$.

This requirement is met by monolayers of hBN and graphene above 80 and 110 °C, respectively (Fig. 2a). Graphene is known to remain stable in humid oxygen atmospheres up to 400 °C (ref. 30), and extrapolation of its conductivity to an operating temperature of 250 °C, at which it is certainly stable, yields extremely high areal conductivities $\sigma > 10^3\,\mathrm{S\,cm}^{-2}$.

Another approach to influencing proton transport through 2D crystals exploits the high affinity of platinum group metals to hydrogen. As shown in Fig. 2b, evaporation of a discontinuous, catalytic layer of Pt or Pd (nominally 1–2 nm thick) onto one of the surfaces of a 2D crystal (see 'Experimental devices' in Methods for fabrication details) resulted in a substantially increased σ. The value of S measured for monolayer hBN became indistinguishable from that of reference bare-hole devices (Fig. 2b), demonstrating that proton permeation (even at room temperature (21–23 °C)) is limited by Nafion's series resistance rather than by passage through the Pt-activated monolayer hBN membrane. Measurements on graphene and bilayer hBN membranes activated with Pt remain little affected by the series resistance and continue to reflect the membranes' intrinsic properties. Temperature-dependent measurements show that Pt reduces the activation energy E by as much as ~0.5 eV (Fig. 2b). This value is in agreement with the ~0.65 eV reduction in E obtained in our simulations of the catalytic effect ('Theoretical analysis of proton transport through 2D crystals' in Methods), which we attribute to attraction of transient protons to Pt (Extended Data Fig. 9). We note that the measurements in Fig. 2b give only a lower limit of ~3 S cm^{-2} for the room-temperature conductivity of catalytically activated monolayer hBN; if this membrane experiences a reduction in E qualitatively similar to that observed for graphene, proton transport across it should be essentially unimpeded.

To demonstrate directly that the applied electric current through our 2D membranes leads to a hydrogen flux, we prepared devices where one of the Nafion–PdH$_x$ contacts is absent and the graphene surface decorated with Pt faces a vacuum chamber equipped with a mass spectrometer (Fig. 3, insets). With either no bias applied between graphene and the remaining PdH$_x$ electrode or a positive bias applied to graphene, we cannot detect any gas leak (including He) between the hydrogen and vacuum chambers (Extended Data Fig. 10). In contrast, applying a negative

bias to graphene causes a steady H$_2$ flux into the vacuum chamber. Its value is determined by the number of protons, I/e (e, elementary charge), passing through the membrane per second. Using the ideal gas law, we find that $F = k_B T(I/2e)$, where the flow rate F is the value measured by the mass spectrometer tuned to molecular hydrogen. The dependence of F on I is shown in Fig. 3 by the solid red line, in excellent agreement with the experiment.

Taken together, our observations establish that monolayers of graphene and hBN constitute a class of proton conductors that raise intriguing questions about the transfer of subatomic particles through atomically thin electron clouds. Moreover, the high proton conductivity, chemical and thermal stability, and impermeability to H$_2$, water and methanol make these membranes attractive candidates for use in various hydrogen technologies. For example, they might be developed into proton membranes for use in fuel cells to solve the problem of fuel crossover and poisoning currently challenging this technology. The demonstrated ability of these membranes to act as a current-controlled source of hydrogen is also appealing for its simplicity and, once large-area graphene and hBN films become commercially available, might be used to extract hydrogen from gas mixtures or air.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Bunch, J. S. et al. Impermeable atomic membranes from graphene sheets. *Nano Lett.* **8,** 2458–2462 (2008).
2. Koenig, S. P., Wang, L., Pellegrino, J. & Bunch, J. S. Selective molecular sieving through porous graphene. *Nature Nanotechnol.* **7,** 728–732 (2012).
3. Leenaerts, O., Partoens, B. & Peeters, F. M. Graphene: a perfect nanoballoon. *Appl. Phys. Lett.* **93,** 193107 (2008).
4. Wang, W. L. & Kaxiras, E. Graphene hydrate: theoretical prediction of a new insulating form of graphene. *New J. Phys.* **12,** 125012 (2010).
5. Miao, M., Nardelli, M. B., Wang, Q. & Liu, Y. First principles study of the permeability of graphene to hydrogen atoms. *Phys. Chem. Chem. Phys.* **15,** 16132–16137 (2013).
6. Tsetseris, L. & Pantelides, S. T. Graphene: an impermeable or selectively permeable membrane for atomic species? *Carbon* **67,** 58–63 (2014).
7. Berry, V. Impermeability of graphene and its applications. *Carbon* **62,** 1–10 (2013).
8. Yuan, W., Chen, J. & Shi, G. Nanoporous graphene materials. *Mater. Today* **17,** 77–85 (2014).
9. Sint, K., Wang, B. & Král, P. Selective ion passage through functionalized graphene nanopores. *J. Am. Chem. Soc.* **130,** 16448–16449 (2008).
10. Jiang, D. E., Cooper, V. R. & Dai, S. Porous graphene as the ultimate membrane for gas separation. *Nano Lett.* **9,** 4019–4024 (2009).
11. Garaj, S. et al. Graphene as a subnanometre trans-electrode membrane. *Nature* **467,** 190–193 (2010).
12. O'Hern, S. C. et al. Selective molecular transport through intrinsic defects in a single layer of CVD graphene. *ACS Nano* **6,** 10130–10138 (2012).
13. Cohen-Tanugi, D. & Grossman, J. C. Water desalination across nanoporous graphene. *Nano Lett.* **12,** 3602–3608 (2012).
14. Kim, H. W. et al. Selective gas transport through few-layered graphene and graphene oxide membranes. *Science* **342,** 91–95 (2013).
15. Li, H. et al. Ultrathin, molecular-sieving graphene oxide membranes for selective hydrogen separation. *Science* **342,** 95–98 (2013).
16. Konatham, D., Yu, J., Ho, T. A. & Striolo, A. Simulation insights for graphene-based water desalination membranes. *Langmuir* **29,** 11884–11897 (2013).
17. Celebi, K. et al. Ultimate permeation across atomically thin porous graphene. *Science* **344,** 289–292 (2014).
18. O'Hern, S. C. et al. Selective ionic transport through tunable subnanometer pores in single-layer graphene membranes. *Nano Lett.* **14,** 1234–1241 (2014).
19. Sun, C. et al. Mechanisms of molecular permeation through nanoporous graphene membranes. *Langmuir* **30,** 675–682 (2014).
20. Stolyarova, E. et al. Observation of graphene bubbles and effective mass transport under graphene films. *Nano Lett.* **9,** 332–337 (2009).
21. Banhart, F., Kotakoski, J. & Krasheninnikov, A. V. Structural defects in graphene. *ACS Nano* **5,** 26–41 (2011).
22. Novoselov, K. S. et al. Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102,** 10451–10453 (2005).
23. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499,** 419–425 (2013).
24. Britnell, L. et al. Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12,** 1707–1710 (2012).
25. Mauritz, K. & Moore, R. B. State of understanding of Nafion. *Chem. Rev.* **104,** 4535–4586 (2004).

Figure 3 | Current-controlled hydrogen flux. Top inset, sketch of our mass spectrometry experiment. Monolayer graphene decorated with Pt nanoparticles separates a vacuum chamber from the Nafion–PdH$_x$ electrode placed under the same H$_2$/H$_2$O conditions as described in 'Conductance measurements' in Methods. Protons penetrate the membrane and recombine into molecular hydrogen. The hydrogen flux (main plot) is detected by a mass spectrometer (Methods). Different symbols refer to different devices. Error bars indicate characteristic fluctuations in the measured signal and the red line is the theoretically expected flow rate. Bottom inset, optical image of one of the devices. Graphene (outlined by the dashed lines) seals a circular aperture 50 μm in diameter etched through the SiN$_x$ membrane (Extended Data Fig. 1). Nafion is underneath the graphene and SiN$_x$ membranes.

26. Morgan, H., Pethig, R. & Stevens, G. T. A proton-injecting technique for the measurement of hydration-dependent protonic conductivity. *J. Phys. E* **19,** 80–82 (1986).
27. Zhong, C. *et al.* A polysaccharide bioprotonic field-effect transistor. *Nature Commun.* **2,** 476 (2011).
28. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS_2 transistors. *Nature Nanotechnol.* **6,** 147–150 (2011).
29. Qiu, H. *et al.* Hopping transport through defect-induced localized states in molybdenum disulphide. *Nature Commun.* **4,** 2642 (2013).
30. Liu, L. *et al.* Graphene oxidation: thickness-dependent etching and strong chemical doping. *Nano Lett.* **8,** 1965–1970 (2008).

METHODS

Experimental devices. Extended Data Fig. 1 explains our microfabrication procedures. We start with preparing free-standing silicon nitride (SiN$_x$) membranes from commercially available Si wafers coated on both sides with 500 nm of SiN$_x$. Reactive ion etching (RIE) is employed to remove a 1×1 mm^2 section from one of the SiN$_x$ layers (steps 1 and 2 in Extended Data Fig. 1). The wafer is then exposed to a KOH solution that etches away Si and leaves a free-standing SiN$_x$ membrane of typically 300×300 μm^2 in size (step 3). During step 4, a circular hole is drilled by RIE through the SiN$_x$ membrane using the same procedures as in steps 1 and 2. Next a 2D crystal (graphene, hBN or MoS$_2$) is prepared by standard micromechanical exfoliation[22] and transferred on top of the membrane using either the wet or dry technique[31,32] to cover the aperture in the SiN$_x$ (step 5). We used hBN crystals commercially supplied by HQ Graphene.

After step 5, the suspended membranes could be examined for their integrity and quality in a scanning electron microscope (SEM). Pristine 2D crystals give little SEM contrast, and it requires some contamination to notice 2D membranes on top of the holes. Contamination can be accidental or induced by the electron beam (Extended Data Fig. 2). If cracks or tears are present, they are clearly seen as darker areas. No such defects could be found in many membranes we visualized in SEM. Occasional cracks such as in Extended Data Fig. 2b were only observed if introduced deliberately or a profound mistake was made during handling procedures. We did not notice any effect of SEM imaging on proton transport but nevertheless avoided prolonged SEM exposures. Because cracks were exceptionally rare, we did not find it necessary to image all the reported devices using SEM.

The fabrication of devices for electrical measurements continues with the deposition of a proton-conducting polymer layer. A Nafion solution (5%, 1,100 equiv. wt) is drop-cast on both sides of a suspended 2D membrane (step 6 in Extended Data Fig. 1). Finally, palladium hydride (PdH$_x$) electrodes are mechanically attached to the Nafion layers. To synthesize these electrodes, a 25 μm-thick Pd foil is left overnight in a saturated hydrogen-donating solution following the procedure of ref. 33. This leads to atomic hydrogen being absorbed into the crystal lattice of Pd, turning it into PdH$_x$. The resulting devices are placed in a water-saturated environment at 130 °C to crosslink the polymer and improve electrical contacts.

The described experimental design is optimized to take into account the following considerations. First, electric currents in Nafion are known to be carried exclusively by protons that hop between immobile sulphonate groups[25]. Nafion is not conductive for electrons, which can be demonstrated directly by, for example, inserting a gold film across a Nafion conductor, which breaks down the electrical connectivity. Accordingly, protons are the only mobile species that can pass between our PdH$_x$ electrodes. Second, PdH$_x$ is widely used as a proton-injecting material that converts an electron (e) flow into a proton (p) one by the following process: PdH$_x \rightarrow$ Pd $+ xp + xe$ (refs 26, 27, 34). This property, combined with the large area of our electrodes (relative to the membrane area A), makes the contact resistance between Nafion and PdH$_x$ negligible such that the circuit conductance in our experiments is limited by either the 2D crystals or, in their absence, the Nafion constriction of diameter D.

For the catalytically activated measurements, 1–2 nm of Pt were deposited by e-beam evaporation directly onto the suspended membrane to form a discontinuous film before the Nafion was deposited. Thicker, continuous films were found to block proton currents. This blocking could be witnessed as the appearance of numerous hydrogen bubbles under the Pt after passage of an electric current. Typically, our Pt films resulted in ~80% area coverage, which reduced the effective area for proton transport accordingly, as found by depositing such Pt films between the Nafion layers, without 2D membranes (see below). Pd was found to be less blocking, and Pd films up to several nanometres thick did not notably impede the proton flow. Otherwise, both Pd and Pt resulted in similar enhancement of proton transport through 2D crystals.

Conductance measurements. The devices described above were placed inside a metal chamber filled with a forming gas (10% H$_2$ in Ar) and containing some liquid water to provide 100% relative humidity. Devices were bonded with gold wires, and I–V curves were recorded using d.c. measurements (Keithley 2636A). We typically varied the voltage in the range of -1 to 1 V at sweep rates up to 0.5 V min^{-1}. We avoided higher voltages because I–V characteristics could become nonlinear and membranes could delaminate as a result of bubble formation. The reported I–V curves were non-hysteretic and highly reproducible. The devices were stable for several weeks if not allowed to dry out.

To characterize our experimental set-up, we first measured leakage currents in the absence of a proton-conductive path. To this end, two metallic contacts were placed on opposite surfaces of a piece of a fresh Si/SiN$_x$ wafer and I–V characteristics were measured under the same humid conditions as above. A conductance of the order of ~5 pS was normally registered. We also used fully processed devices and then mechanically removed the Nafion film and electrodes. In the latter case, the parasitic conductance was slightly (a factor of two) higher, which is probably due

to a processing and polymer residue left on SiN$_x$. In principle, it would be possible to reduce the leakage currents by using, for example, separate chambers on opposite sides of the Si wafer[11], but the observed parasitic conductance was deemed small enough for the purpose of the present work.

As a reference, we studied the conductivity of bare-hole devices that were prepared in exactly the same manner as our membrane devices but without a 2D crystal covering the aperture (step 5 in Extended Data Fig. 1 was omitted). Extended Data Fig. 3a shows the conductance of such devices as a function of their diameter D. Within the experimental scatter, conductance S increases linearly with D, in agreement with Maxwell's formula[35]: $S = \sigma_N D$. The latter is derived by solving Laplace's equation for two semi-spaces that have conductivity σ_N and are connected by a tube with D much larger than its length d. In our case, $d = 500$ nm and the condition is satisfied, except possibly by the smallest membranes with $D = 2$ μm.

From the dependence shown in Extended Data Fig. 3a, we can estimate the bulk conductivity of our Nafion films as ~1 mS cm^{-1}. As shown in the main text, Nafion's conductivity did not limit our measurements of proton transport through 2D crystals, except for the case of catalytically activated monolayer hBN. Nonetheless, we note that the found σ_N value is two orders of magnitude smaller than values achievable for highest-quality Nafion[36]. There are two reasons for this. First, solution-cast Nafion like that used in our experiments is known to be typically one order of magnitude lower in conductivity than the highest-quality Nafion[37,38]. Second, to achieve the highest conductivity, Nafion is normally pre-treated by boiling in H$_2$O$_2$ and H$_2$SO$_4$ for several hours[36-38]. When this procedure was used, our Nafion films indeed increased their conductivity by a factor of ten, reaching the standard values for solution-cast Nafion of ~10 mS cm^{-1}. Unfortunately, this harsh treatment destroyed our membrane devices, with the Nafion delaminating from SiN$_x$, and so could not be used. Proton concentrations can be estimated[39] from σ_N and, for our films, are expected to be ~0.1 M.

For consistency, most of the 2D membranes reported in the main text were 2 μm in diameter. However, we studied many other membranes with D ranging from 1 to 50 μm. Their conductances are found to scale linearly with the aperture area A. Extended Data Fig. 3b shows this for ten monolayer hBN devices with D between 1 and 4 μm. Within the experimental scatter for devices of the same D, the conductance increases linearly with A, in agreement with general expectations. The same scaling was also observed for graphene membranes.

Reproducibility. Figures 1b and 2b show that our measurements of σ were highly reproducible for different devices of nominally the same size. The scatter in σ can be attributed to accidental contamination that blocks proton currents through parts of the 2D membranes. Further evidence of little variation in σ for different devices is provided by the correct scaling of σ with membrane area (Extended Data Fig. 3b). It is important to emphasize that, because of the exponential dependence of σ on T, the high reproducibility of σ at room temperature implies that the activation energies E also cannot differ much for different devices. Nonetheless, to show directly that E is device independent, Extended Data Fig. 4 plots $\sigma(T)$ for three bilayer hBN membranes. The best fits respectively yield $E = 0.65, 0.59$ and 0.57 eV. These values fall within the uncertainty interval (0.61 ± 0.04 eV) stated for bilayer hBN in the main text. Furthermore, the inset of Extended Data Fig. 4 compares $\sigma(T)$ for the device shown in Fig. 2a with data obtained for three other graphene membranes. These devices failed during measurements, presumably owing to mechanical strain induced by changes in T. However, the data acquired before the devices broke show that all the membranes have the same activation energy.

Although Nafion was the material of choice in this work owing to its stability and convenience of handling, to prove the generality of our results we also investigated the proton conductivity of 2D crystals when they were immersed in water. For these experiments, 2D membranes were fabricated in the same way as described previously, but, instead of covering the 2D crystals with Nafion, they were used to separate two reservoirs containing liquid electrolytes (Extended Data Fig. 5). Typical I–V characteristics recorded for membranes made from mono-, bi-, and tri-layer hBN in the liquid-cell set-up are presented in Extended Data Fig. 5a. They were recorded using chronoamperometry, and the values shown in the figure correspond to stable currents. The current response was symmetric for positive and negative biases. For devices prepared in the same manner but without a 2D membrane, the conductance S was $>10^3$ times higher than in the presence of monolayer hBN, which ensured that the 2D crystals limited the proton current in the liquid-cell set-up. As in the case of Nafion, we also found a parasitic parallel conductance, but it was somewhat higher (~20 pS) because of the liquid environment. Although it should be possible[11] to reduce the leakage current in the liquid-cell set-up, we find the present accuracy sufficient for our objectives. Within this uncertainty, we could not detect any proton current through either trilayer hBN or, as for the Nafion set-up, monolayer MoS$_2$, bilayer graphene or any thicker 2D crystals. The observed proton conductivity was highly reproducible for different devices, as shown by the statistics in Extended Data Fig. 5b. Most importantly, the measured proton conductivities agree well with the values found using Nafion as the proton-conducting

medium (compare Fig. 1b with Extended Data Fig. 5b). We note that the devices used in the liquid-cell experiments were more fragile than those in the Nafion experiments and survived for shorter times because of the lack of mechanical support. Accordingly, we focused in our present work on Nafion devices.

Absence of atomic-scale defects. As discussed above, visual inspection of membranes using SEM can reliably rule out holes and cracks with sizes down to $\sim$10 nm (Extended Data Fig. 2b). However, SEM cannot resolve nanometre- or atomic-scale defects, and other techniques are necessary to rule out pinholes of these sizes. As already mentioned in the main text, no such defects have ever been reported for pristine graphene obtained by micromechanical cleavage in numerous transmission electron microscopy and scanning tunnelling microscopy studies over many years. To add to this argument in the case of our particular membranes, we used Raman spectroscopy, which is known to be extremely sensitive to atomic-scale defects in graphene. The intensity of the D peak ($\sim$1,350 cm^{-1}) provides a good estimate of their concentration. Importantly, atomic-scale defects can be not only vacancies or larger pinholes but also adatoms that should not allow protons through. Therefore, the D peak provides the upper limit on the concentration of pinholes. Despite our efforts, we could not discern any D peak in our graphene membranes[40]. These measurements set a limit on possible pinhole densities as $\sim$10^8 cm^{-2}, or one defect per μm^2 (ref. 40). Furthermore, such a low density of defects in graphene (including adatoms) is in stark contrast with a high density ($\sim$10^{13} cm^{-2}) of sulphur vacancies found in mechanically cleaved MoS$_2$ (ref. 29). Nevertheless, no proton current could be detected in our MoS$_2$ membranes. If we assume each vacancy to provide a hole of $\sim$1 Å in size, the expected $\sim$10^5 vacancies present in our typical MoS$_2$ membranes would provide an effective opening $\sim$30 nm in diameter. Using the results of Extended Data Fig. 3a, this is expected to lead to a conductance of $\sim$3 nS, which is >100 times larger than the limit on σ set by our measurements for monolayer MoS$_2$. This indicates that individual vacancies may increase the proton conductance much less than their classical diameter suggests. This conclusion is confirmed by using devices made from graphene and hBN monolayers, which were grown by chemical vapour deposition (CVD). Such CVD materials are known to contain many atomic-scale defects, as evidenced, for example, by a strong D peak. Nevertheless, CVD membranes had the same proton conductivity as that found for cleaved monolayers. This unambiguously shows that, even if a few atomic-scale pinholes were present in cleaved 2D crystals, they could not noticeably contribute to the reported σ.

To strengthen the above arguments further, we tried to rule out the presence of even individual vacancies in our cleaved graphene and hBN devices. The most sensitive technique known to detect pinholes is arguably the measurement of gas leakage from small pressurized volumes[1,2]. To this end, a microcavity typically $\sim$1 μm^3 in size is etched in a Si/SiO$_2$ wafer, sealed with graphene or hBN and then pressurized. If the pressure inside the microcavity is higher than that outside, the membrane bulges upwards; if it is lower, downwards. Changes in pressure can be monitored by measuring the height of the bulge as a function of time using atomic force microscopy (AFM). If there are no holes in the membrane, the gas leaks out slowly along the SiO$_2$ layer: it typically takes many hours until the pressures inside and outside the microcavity equalize[1]. However, the presence of even a single atomic-scale hole, through which atoms can effuse, allows the pressure to equalize in less than a second[2]. Following the procedures reported previously[1,2], we prepared microcavities in a Si/SiO$_2$ wafer and sealed them with cleaved monolayer graphene. The microcavities were placed inside a chamber filled with Ar at 200 kPa for typically four days to gradually pressurize them. After taking the devices out, the membranes were found to bulge upwards. Extended Data Fig. 7 shows the deflation of such microballoons with time. In agreement with the previous report[1], the Ar leak rates were found to be $\sim$10^3 atoms per second. If one or a few atomic-scale holes were introduced by, for example, ultraviolet chemical etching, the leak rate increased by many orders of magnitude, leading to practically instantaneous deflation[2]. This shows again that no atomic-scale defects were present in our membranes obtained by mechanical cleavage.

Nafion-limited conductivity. We have reported in the main text that the proton conductivity of catalytically activated monolayer hBN is so high that the series resistance of Nafion becomes the limiting factor in our measurements. This observation is further illustrated in Extended Data Fig. 8 by comparing T dependences for different devices in which Nafion was the limiting factor (bare-hole, Nafion/Pt/Nafion and hBN-with-Pt devices). Consistent with the small activation energy for proton transport in Nafion (<0.02 eV; ref. 36), we found that temperature effects in all such devices are small over the entire T range (Extended Data Fig. 8). The non-monotonic T dependence for the devices with a Pt layer remains to be understood, but we note that Nafion often exhibits similar non-monotonic behaviour[41] at higher T, beyond the temperature range shown in Extended Data Fig. 8. We speculate that the Pt activation shifts this peak to lower T. Importantly, the influence of Pt on local conductivity in the Nafion constriction is approximately the same independently of whether or not an hBN membrane is present. This confirms that the proton conductivity of

Pt-activated hBN is so high that it becomes unmeasurable in our experiments. It would require membranes with much larger D to determine σ for catalytically activated hBN.

Theoretical analysis of proton transport through 2D crystals. It is possible to understand the differences that we find in σ by considering the electron clouds created by different 2D crystals. These clouds impede the passage of protons through 2D membranes. In addition to the plots of the electron density in Fig. 1b, Extended Data Fig. 6 shows similar plots of the electron clouds with superimposed positions of C, B and N atoms using the ball-and-stick model of the graphene and hBN crystal lattices. In addition, Extended Data Fig. 6 plots the electron density for monolayer MoS$_2$ consisting of a monolayer of Mo atoms sandwiched between two monolayers of sulphur. One can immediately see that the latter cloud is much denser than those of monolayer hBN and graphene, which qualitatively explains the absence of proton transport through MoS$_2$ monolayers.

For quantitative analysis, let us first note that proton permeation through graphene has previously been studied[4–6] using both *ab initio* molecular dynamics (AIMD) simulations and the climbing-image nudged elastic band (CI-NEB) method. These studies have provided estimates for the proton barrier E created by graphene, which range from $\sim$1.17 to 2.21 eV (refs 4–6). We reproduced those results for the case of graphene and extended them to monolayer hBN. Our simulations were performed using the CP2K package[42] with the Pade exchange correlation functional form[43]. The energy cut-off of plane-wave expansions was 380 Ry, and we used the double-ζ valence basis with one set of polarization functions[44] and the Goedecker–Teter–Hutter pseudopotentials[43]. In the first approach, the bombardment of graphene and monolayer hBN with protons of varying kinetic energy was simulated using AIMD in the NVE ensemble (that is, the number of atoms, the volume and the energy are assumed to be constant). The barrier was estimated to be the minimum kinetic energy necessary for proton transfer. The AIMD simulations have yielded E for graphene of between 1.30 and 1.40 eV, in good agreement with refs 4, 5.

In the second (CI-NEB) approach, we calculated the energy for various configurations (usually referred to as 'images'), which correspond to different distances between a proton and a 2D membrane[45]. This provided a series of images for a proton approaching the membrane. The energy was then minimized over obtained images and plotted as a function of proton–crystal distance. The barrier E was estimated using the differential height of such energy profiles. Extended Data Fig. 9 shows examples of these profiles for graphene and monolayer hBN. From the CI-NEB calculations, we estimate the proton barriers to be 1.26 and 0.68 eV for graphene and monolayer hBN, respectively, in agreement with our AIMD results. Finally, to model the effect of Pt on proton transport, we again used AIMD simulations. To this end, four Pt atoms were placed at a fixed distance of 4 Å from the graphene membrane and the bombardment with protons was simulated as described above (Extended Data Figs 9c, d). The addition of the Pt atoms resulted in a significant reduction of the graphene barrier to $\sim$0.6 eV; that is, by a factor of two. The absolute value of the reduction in the barrier height is in agreement with the experiment.

Our measurements also show that I–V characteristics remain linear over a wide range of biases V (up to 1.5 V in the case of Extended Data Fig. 5a). This observation is surprising because the voltage drop across the proton barrier becomes comparable to the barrier height divided by the charge of proton, E/e. Under these circumstances, one intuitively expects a considerable increase in the barrier transparency and strongly nonlinear I–V characteristics, as happens in the case of electron tunnelling. To understand the observed linear behaviour, we modelled our experimental situation using both AIMD and CI-NEB simulations. Additional accelerating fields of up to 1 V nm^{-1} were applied across a graphene sheet. We have found that E changes little, by only $\sim$15 meV for the highest simulated field. Because of inevitable screening by mobile ions, we expect significantly lower electric fields in our experiments than 1 V nm^{-1}, which implies that E changes by much less than $k_\mathrm{B}T$. The low sensitivity of E with respect to V is in agreement with the linear I–V characteristics observed experimentally, but the physical origin of this behaviour remains to be understood. Tentatively, we attribute it to the following: applied voltage not only accelerates protons but also polarizes the electron clouds of 2D crystals, which in turn leads to significant deceleration of protons.

Detection of proton flow by mass spectrometry. To illustrate that the electric current through our 2D membranes is carried by hydrogen ions, we used an alternative set-up described in the main text and shown in more detail in Extended Data Fig. 10a. Protons transferring through graphene are collected at a catalyst Pt layer where they recombine to form molecular hydrogen: $2p + 2e \rightarrow H_2$. The hydrogen flux is then measured with a mass spectrometer (Inficon UL200). Because the electric current I is defined by the number of protons passing through the graphene membrane, the hydrogen flow F is directly related to the passing current I, with no fitting parameters (see the main text).

For this particular experiment, the membrane devices were made as large as possible (50 μm in diameter) to increase the hydrogen flux to values that could be detected using our mass spectrometer. To collect the electric current at the graphene

membrane, a metallic contact (100 nm Au/5 nm Cr) was fabricated next to the SiN$_x$ aperture, before graphene was transferred on top, to cover both aperture and contact (right inset of Fig. 3). This side of the Si wafer (with graphene on top) was then decorated with 1–2 nm of Pt to increase the proton flux. The opposite face of the graphene membrane was covered with Nafion and connected to a PdH$_x$ electrode in the way described above. The resulting device on the Si wafer was glued with epoxy to a perforated Cu foil that was clamped between two O rings to separate two chambers: one filled with a gas and the other connected to the mass spectrometer (Extended Data Fig. 10a). First, we always checked for possible leaks by filling the gas chamber with helium at atmospheric pressure. No He leak could be detected above background readings of the spectrometer ($\sim 10^{-8}$ bar cm^3 s^{-1}). Then the chamber was filled with our standard gas mixture (10% H$_2$ in Ar at 1 bar and at 100% humidity). No hydrogen flux could be detected without applying negative bias to the graphene. By applying such a bias a controllable flow of H$_2$ at a level of $\sim 10^{-5}$ bar cm^3 s^{-1} or $\sim 10^{14}$ hydrogen molecules per second was readily detected (Extended Data Fig. 10b). This figure shows the hydrogen flow rates F as a function of time for one of our devices using negative biases from 0 to 20 V. When cycling back from 20 to 0 V, the curves retraced themselves, indicating that the membrane was undamaged during the measurements.

Atomic hydrogen is highly unstable with respect to its molecular form, and it is most likely that the conversion into molecular hydrogen takes places at the surface of Pt rather than in the vacuum chamber. Accordingly, the Pt layer has to be discontinuous to let hydrogen escape. For continuous coverage ($>$5 nm of Pt), we observed formation of small hydrogen bubbles that grew as we increased the amount of electric charge passed through the circuit. The largest bubbles eventually burst. It is also instructive to mention the case in which a continuous Au films was evaporated on top of the above devices (already containing a discontinuous Pt layer). We found that a bias applied across such devices resulted in the formation of large bubbles at the interface between the graphene and the metal film. The bubbles could burst and sometimes damaged the membrane. This precluded the use of continuous metal films for the mass spectrometry experiment. The same bubbling effect was observed for hBN membranes covered with a Pt film providing continuity of the electrical circuit for insulating hBN. These observations serve as yet another indication of proton transfer through graphene and hBN membranes. However, no bubbles could be observed for thicker 2D crystals, which again proves their impermeability to protons.

31. Nair, R. R. *et al.* Graphene as a transparent conductive support for studying biological molecules by transmission electron microscopy. *Appl. Phys. Lett.* **97,** 153102 (2010).
32. Kretinin, A. V. *et al.* Electronic properties of graphene encapsulated with different two-dimensional atomic crystals. *Nano Lett.* **14,** 3270–3276 (2014).
33. Murphy, D. W. *et al.* A new route to metal hydrides. *Chem. Mater.* **5,** 767–769 (1993).
34. Schuldiner, S., Castellan, G. W. & Hoare, J. P. Electrochemical behavior of the palladium-hydrogen system. I. Potential-determining mechanisms. *J. Chem. Phys.* **28,** 16–19 (1958).
35. Agraït, N., Yeyati, A. L. & van Ruitenbeek, J. M. Quantum properties of atomic-sized conductors. *Phys. Rep.* **377,** 81–279 (2003).
36. Sone, Y., Ekdunge, P. & Simonsson, D. Proton conductivity of Nafion 117 as measured by a four electrode AC impedance method. *J. Electrochem. Soc.* **143,** 1254–1259 (1996).
37. Silva, R., De Francesco, M. & Pozio, A. Solution-cast Nafion® ionomer membranes: preparation and characterization. *Electrochim. Acta* **49,** 3211–3219 (2004).
38. Deluca, N. & Elabd, Y. Nafion/poly(vinyl alcohol) blends: effect of composition and annealing temperature on transport properties. *J. Membr. Sci.* **282,** 217–224 (2006).
39. Spry, D. B. & Fayer, M. D. Proton transfer and proton concentrations in protonated Nafion fuel cell membranes. *J. Phys. Chem. B* **113,** 10210–10221 (2009).
40. Alberti, G., Casciola, M., Massinelli, L. & Bauer, B. Polymeric proton conducting membranes for medium temperature fuel cells (110–160 °C). *J. Membr. Sci.* **185,** 73–81 (2001).
41. Ni, Z. H. *et al.* On resonant scatterers as a factor limiting carrier mobility in graphene. *Nano Lett.* **10,** 3868–3872 (2010).
42. VandeVondele, J. *et al.* Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167,** 103–128 (2005).
43. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54,** 1703–1710 (1996).
44. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127,** 114105 (2007).
45. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113,** 9901–9904 (2000).
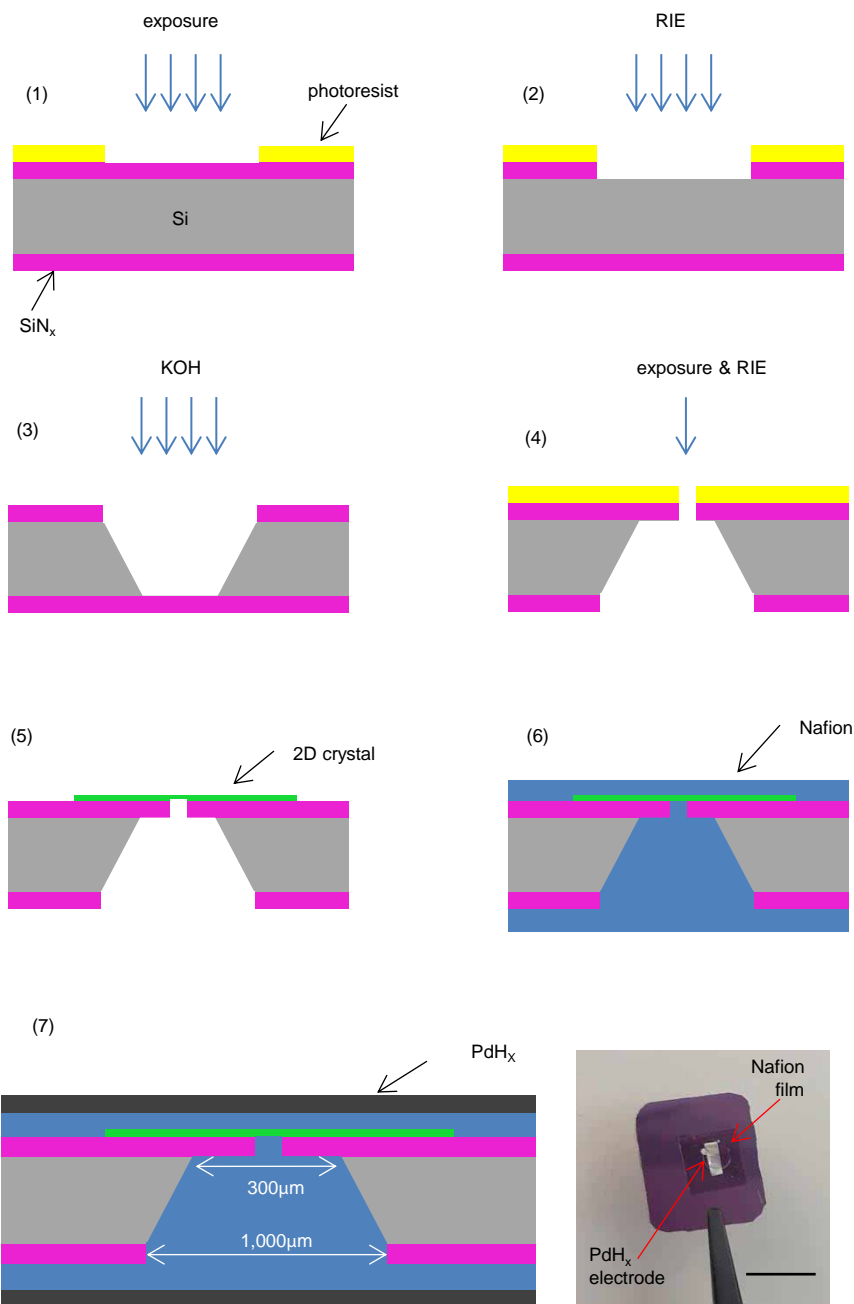
Extended Data Figure 1 | Microfabrication process flow. (1) An etch mask is made by photolithography. (2) RIE is used to remove the exposed SiN_x layer. (3) Si underneath is etched away by wet chemistry. (4) By repeating steps 1 and 2, a hole is drilled through the membrane. (5) The 2D crystal is transferred to cover the etched hole. (6) Nafion is deposited on both sides of the wafer. (7) PdH_x electrodes are attached. Bottom right, optical photo of the final device. Scale bar, 1 cm.

a

b
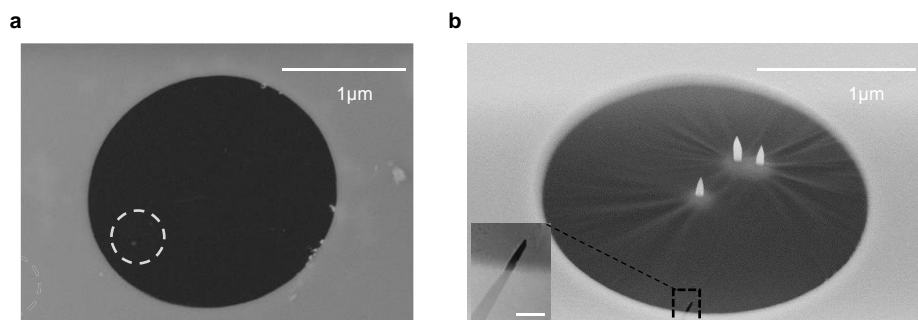
Extended Data Figure 2 | SEM images of suspended 2D crystals.
a, Monolayer graphene with some accidental contamination. One of the particles away from the edge is marked with a white circle. **b**, Suspended graphene with pillars of hydrocarbon contamination intentionally induced by a focused electron beam. The inset shows a crack in the membrane; scale bar, 100 nm.

a

b
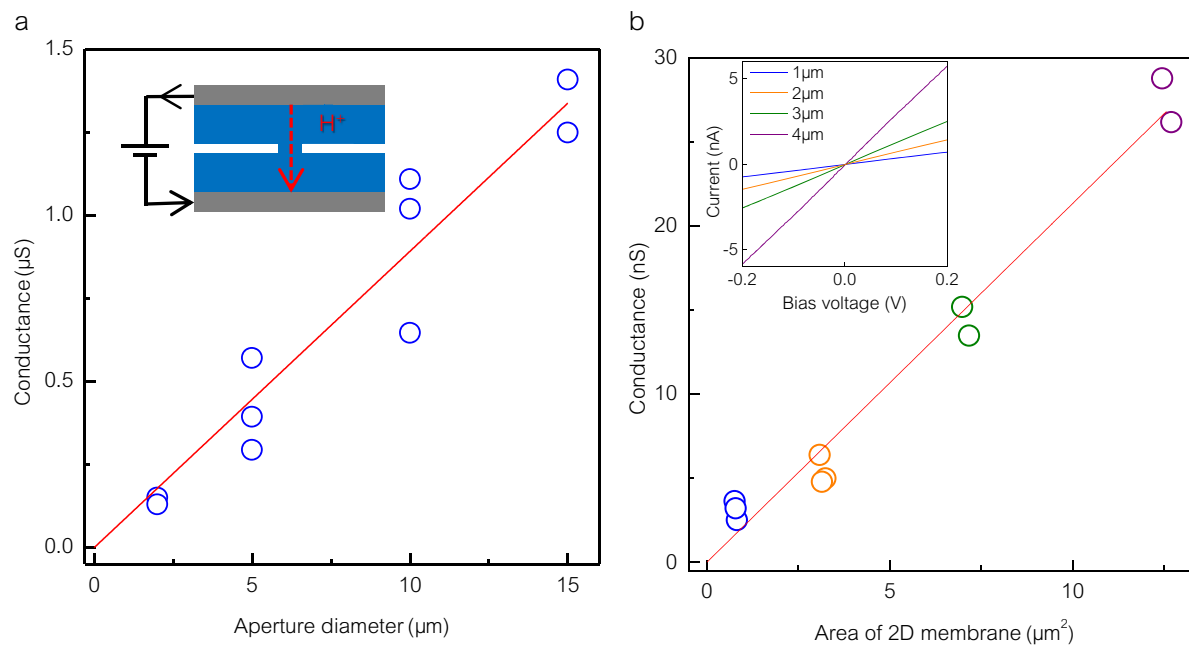
Extended Data Figure 3 | Dependence of proton conductance on aperture size. a, A bare-hole device exhibits a linear dependence of σ on the aperture diameter, as expected for this geometry[35]. The inset is a sketch of such a device. **b**, Proton conductance through monolayer hBN scales quadratically with membrane diameter, that is, linearly with membrane area. The inset shows examples of I–V characteristics for four hBN monolayer devices with different D values, from 1 to 4 µm.
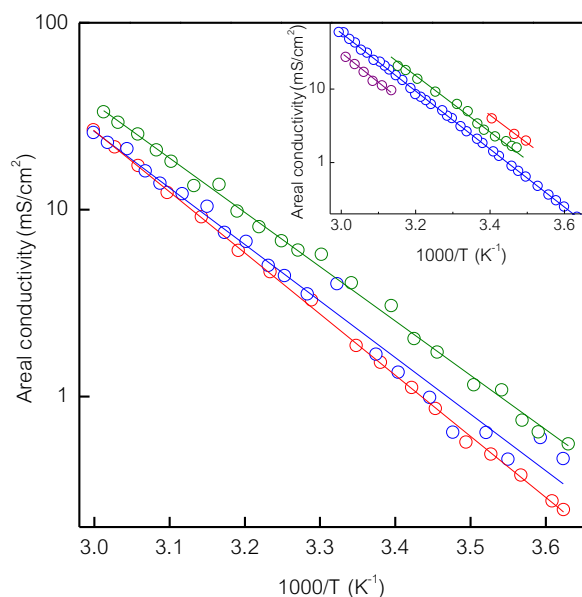
Extended Data Figure 4 | Reproducibility of proton barrier heights for different devices. Activation temperature dependences for three bilayer hBN devices (symbols are the experimental data; lines are the best fits). Inset: equivalent data for four monolayer graphene devices, three of which could be measured only within limited T intervals before they failed. The blue line is the best fit to the Arrhenius-type dependence; the other lines are guides to the eye indicating that all the devices exhibit practically the same E.

a

b
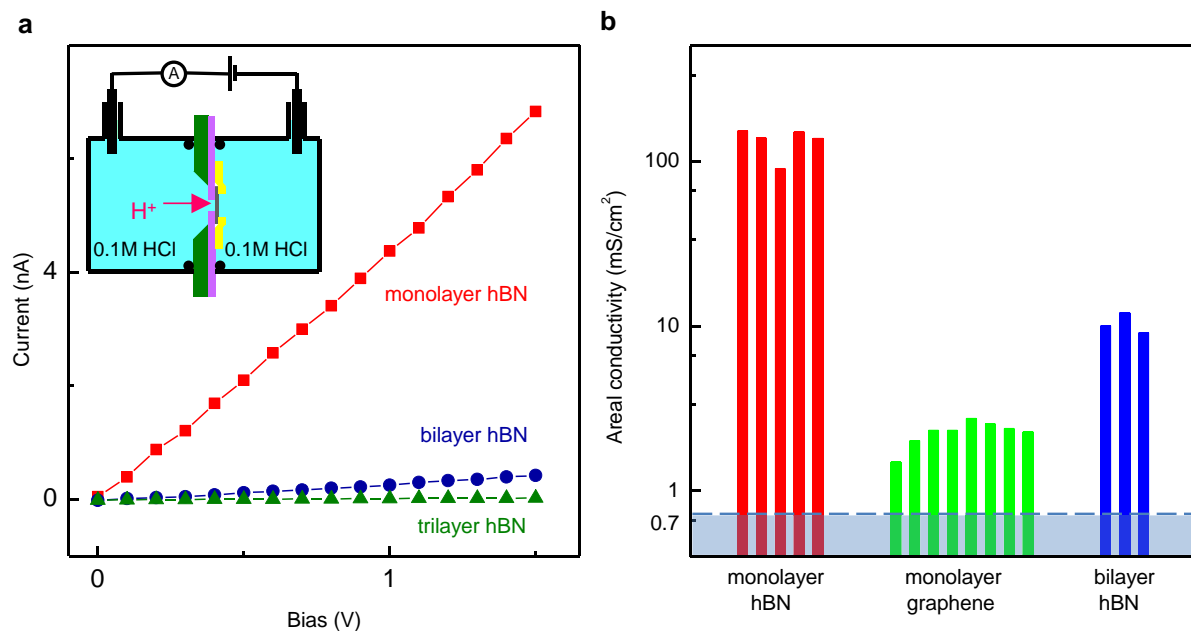
Extended Data Figure 5 | Proton transport through 2D crystals in electrolytes. a, Examples of *I–V* characteristics for mono-, bi- and trilayer hBN membranes covering an aperture 2 μm in diameter. The inset shows a sketch of the liquid-cell set-up. To match the proton concentration in our Nafion experiments, we used a 0.1 M HCl solution in both containers. An additional polymer seal (yellow) is used to avoid leakage along the 2D crystal/substrate interface[11]. Ag/AgCl electrodes are placed inside each reservoir to measure ionic currents. In the case of trilayer hBN, the measured current falls within the range given by leakage currents. **b**, Histograms for the 2D crystals that exhibited unambiguous proton conductivity in the liquid-cell set-up. Each bar represents a different 2 μm membrane. The shaded area shows our detection limit.
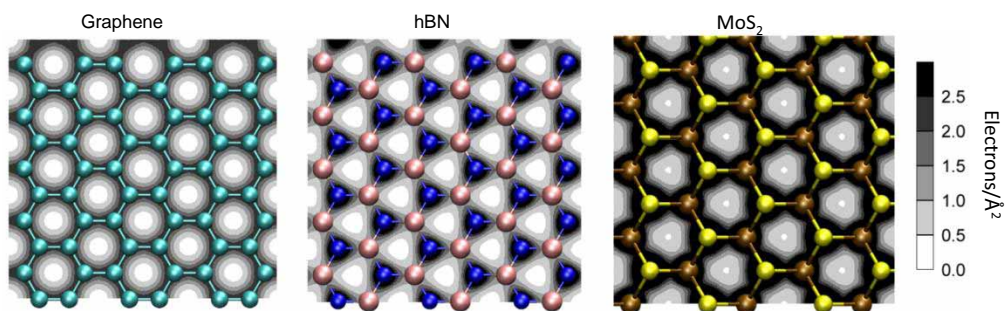
Extended Data Figure 6 | Electron clouds of 2D crystals. Integrated charge densities for graphene, monolayer hBN (nitrogen is indicated by blue balls; boron in pink) and monolayer MoS_2 (S is in yellow; Mo in brown).

a

b
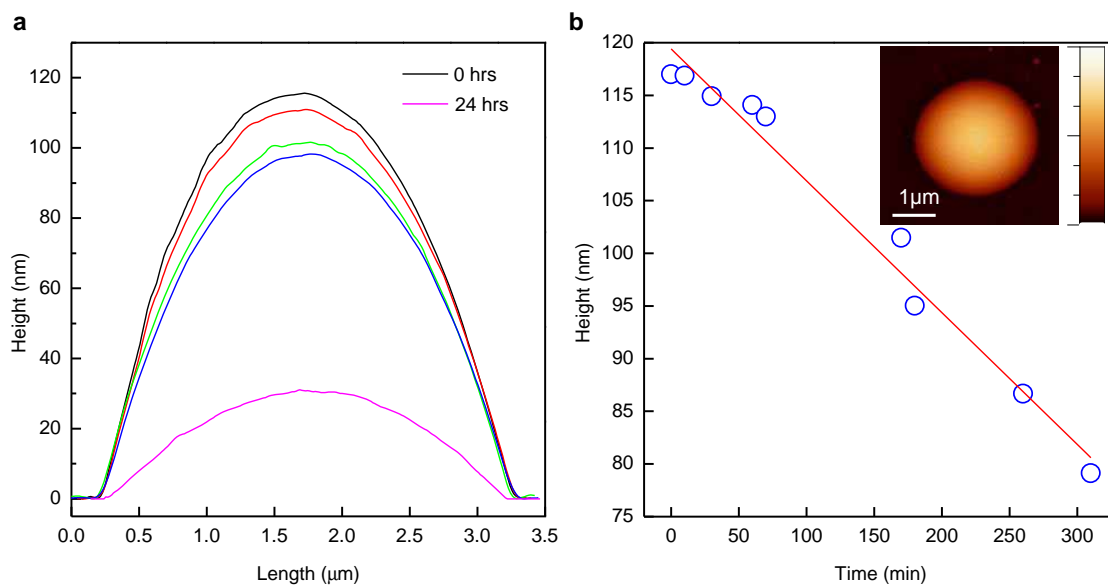
Extended Data Figure 7 | Slow deflation of micro-balloons rules out atomic-scale pinholes. a, Height profiles for a typical graphene membrane over 24 h of observation. **b**, Maximum height as a function of time. The inset shows a typical AFM image of a pressurized graphene microcavity (colour scale, 0–130 nm). We measured six graphene membranes and all of them showed the same deflation rates, independently of whether or not Pt was deposited on top. Similar behaviour was observed for hBN monolayers.
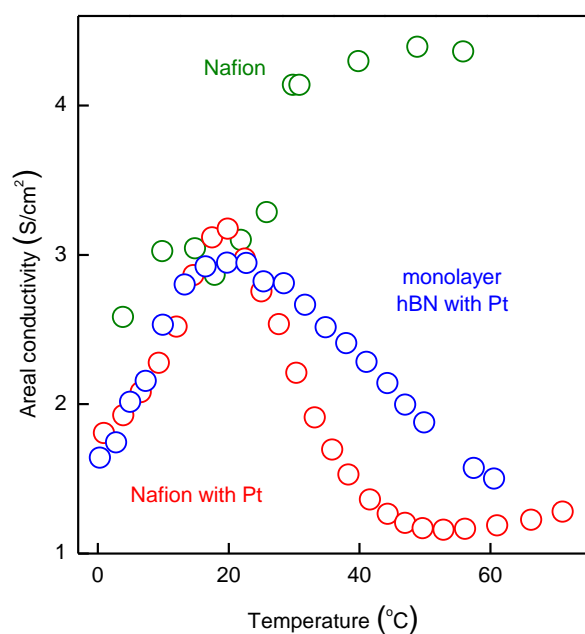
Extended Data Figure 8 | Nafion-limited conductivity for Pt-activated hBN.
Temperature dependences for a bare-hole device (constriction with Nafion only), a Nafion/Pt/Nafion device (no 2D membrane present) and a membrane device with catalytically activated monolayer hBN. The nominal conductivity is calculated as the measured conductance S divided by the aperture area A.
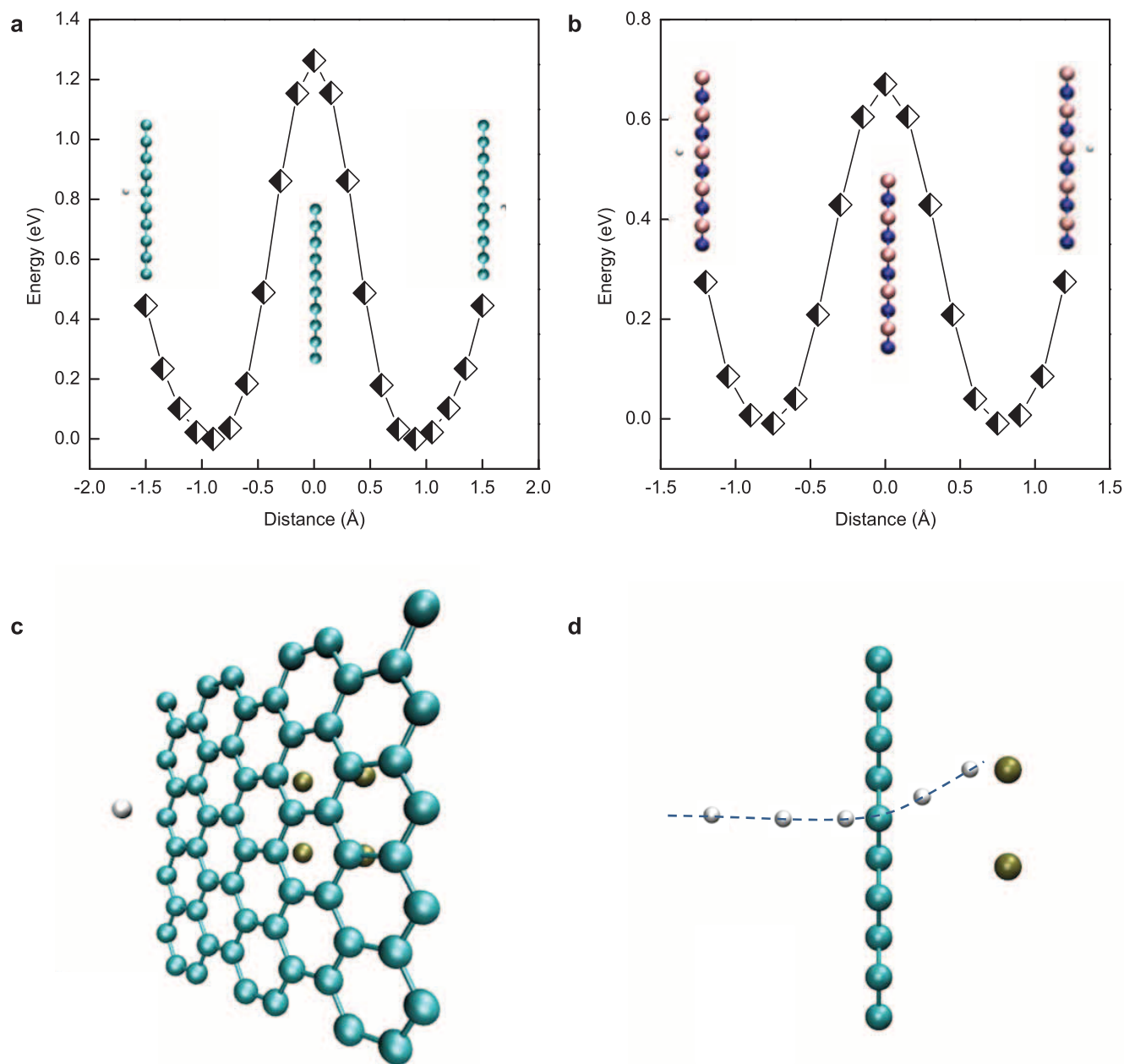
Extended Data Figure 9 | Simulations of proton transport through 2D crystals. a, b, Profiles of energy as a function of the distance of the proton to the centre of the hexagonal ring in graphene (**a**) and hBN (**b**), calculated using the CI-NEB method. Carbon atoms are shown in cyan, nitrogen in blue, boron in pink and protons in white. **c**, The influence of catalytic nanoparticles used in the experiment is mimicked by placing four Pt atoms at a distance of 4 Å from the graphene sheet. **d**, Trajectory of protons with an initial kinetic energy of 0.7 eV (the other two Pt atoms cannot be seen because of the perspective). The bent trajectory indicates that the decrease in barrier height is due to interaction of protons with Pt. Carbon atoms are shown in cyan, Pt in ochre and protons in white.

a

b
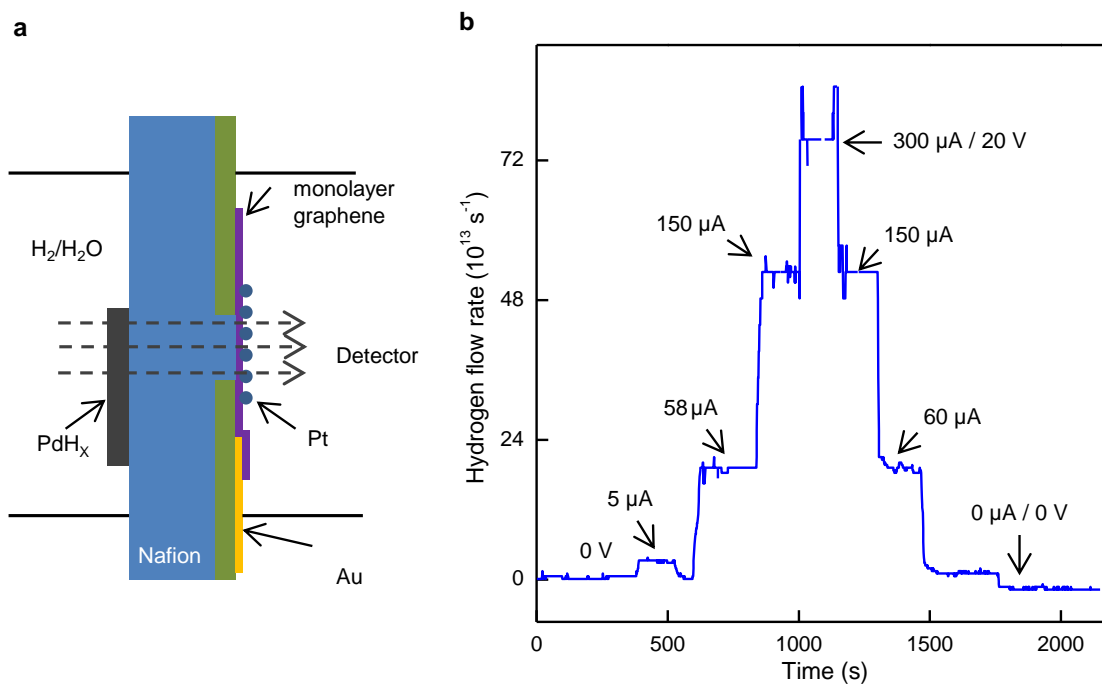
Extended Data Figure 10 | Hydrogen flow detection. a, Schematics of our devices for mass spectroscopy measurements. **b**, Example of the observed hydrogen flow rates as a functions of time for different negative biases on the graphene membrane. The voltage was applied in steps and resulted in the current values indicated next to the steps.

LETTER

Formation and properties of ice XVI obtained by emptying a type sII clathrate hydrate

Andrzej Falenty[1], Thomas C. Hansen[2] & Werner F. Kuhs[1]

Gas hydrates are ice-like solids, in which guest molecules or atoms are trapped inside cages formed within a crystalline host framework (clathrate) of hydrogen-bonded water molecules[1]. They are naturally present in large quantities on the deep ocean floor and as permafrost, can form in and block gas pipelines, and are thought to occur widely on Earth and beyond. A natural point of reference for this large and ubiquitous family of inclusion compounds is the empty hydrate lattice[1–6], which is usually regarded as experimentally inaccessible because the guest species stabilize the host framework. However, it has been suggested that sufficiently small guests may be removed to leave behind metastable empty clathrates[7,8], and guest-free Si- and Ge-clathrates have indeed been obtained[9,10]. Here we show that this strategy can also be applied to water-based clathrates: five days of continuous vacuum pumping on small particles of neon hydrate (of structure sII) removes all guests, allowing us to determine the crystal structure, thermal expansivity and limit of metastability of the empty hydrate. It is the seventeenth experimentally established crystalline ice phase[11], ice XVI according to the current ice nomenclature, has a density of 0.81 grams per cubic centimetre (making it the least dense of all known crystalline water phases) and is expected[7,12] to be the stable low-temperature phase of water at negative pressures (that is, under tension). We find that the empty hydrate structure exhibits negative thermal expansion below about 55 kelvin, and that it is mechanically more stable and has at low temperatures larger lattice constants than the filled hydrate. These observations attest to the importance of kinetic effects and host–guest interactions in clathrate hydrates, with further characterization of the empty hydrate expected to improve our understanding of the structure, properties and behaviour of these unique materials.

The two main gas hydrate structure types[1], sI and sII, both have a cubic symmetry. Topologically, they are related to the SiO_2 phases of melanophlogite and dodecasil-3C, respectively[13], and to the Si- and Ge-clathrates that have been obtained in a guest-free form[9,10]. Because the direct nucleation of an empty hydrate lattice from liquid water under tension (negative pressure) is challenging if not impossible, we followed the previously suggested[7,8,10] approach of pumping on a clathrate with small guest molecules to remove them. This can be a very slow process with guest molecules like CH_4 or CO_2, which cannot pass through the 5- and 6-membered hydrogen-bonded water rings present in the clathrate without the presence of a water vacancy (a 'hole-in-the-cage')[14] (see Methods). But there is evidence that smaller guests, like H_2, may diffuse through the lattice without such vacancy-mediated assistance[15,16], with similar behaviour expected for He and Ne, which can also enter the open hexagonal channels of the ice Ih structure[17]. For this reason and because we found a fast and convenient way to produce large amounts of it, we used as starting material deuterated Ne clathrate (Fig. 1; see Methods for production details).

Samples of Ne clathrate were pumped at constant temperatures between 110 and 145 K for several hours while neutron diffraction data were taken. The cage occupancies were obtained from full-pattern Rietveld refinements as a function of pumping time, indicating a progressive emptying of the cages. Emptying proceeded distinctly slower for the small 5^{12} cages (SC) than for the large $5^{12}6^4$ cages (LC) (Extended Data Fig. 1). The empty hydrate structure decomposes at temperatures of 145 K and above, much like the sII H_2 clathrate[18] where exposure of the sample to a reduced pressure presumably causes uncontrolled H_2 out-diffusion and thereby reduced cage filling. The final pumping attempt was therefore conducted at 142 K and run for 5 days, after which the cages were found to be empty within the 2σ limit of precision. A neutron diffraction structure analysis of this empty hydrate sample at 5 K, to probe details of its hydrogen-bonded water topology, yielded mean atomic coordinates of the disordered oxygen and deuterium positions (Table 1). A neutron diffraction study of the initial Ne clathrate under identical conditions was also carried out to compare structures (see Extended Data Fig. 2). Finally, empty hydrate and Ne clathrate samples were heated in steps of 10 K up to 140 K to study lattice constants and thermal expansivity (see Methods).

The empty and the Ne-filled clathrate are both proton-disordered (or rather, deuteron-disordered in the present case) due to orientational disorder of their constituent water molecules, and have similar overall values of the time–space averaged hydrogen-bond distances of 2.751 and 2.748 Å, respectively. But individual hydrogen-bond distances differ significantly, which translates into relative differences in the volumes of SC and LC: while the small cage expands by 3.9‰ from 159.57(9) to 160.2(1) Å^3 upon Ne removal, the large cage expands by 3.3‰ from 306.3(2) to 307.3(3) Å^3. The lattice constants and expansivities of the two structures also show significant differences, with the Ne clathrate having the smaller lattice constant at low temperature and the larger expansivity over the investigated temperature range (Fig. 2). In clear
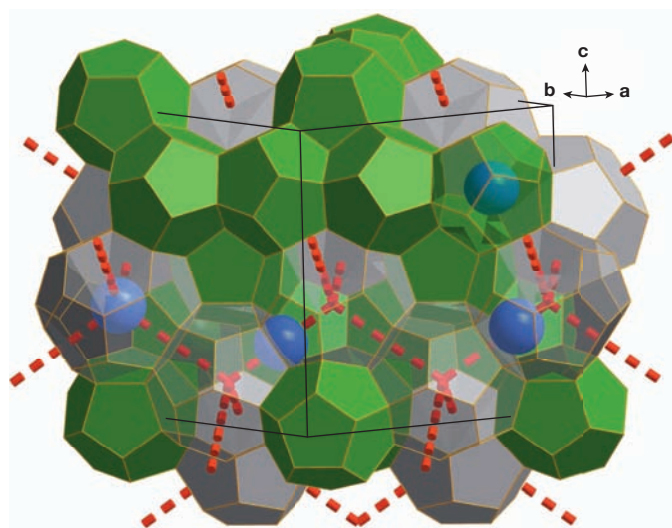
Figure 1 | Leaching of Ne atoms from the sII clathrate structure. Ne atoms (in blue) can easily travel between large cages (in grey) passing through six-membered rings of water molecules (red dashed lines). Removal of Ne atoms from the small cages (in green) requires the presence of a water vacancy in one of the five-membered rings[14].

[1]GZG Abteilung Kristallographie, Universität Göttingen, Goldschmidtstrasse 1, 37077 Göttingen, Germany. [2]Institut Laue-Langevin (ILL), 71 avenue des Martyrs, 38000 Grenoble, France.

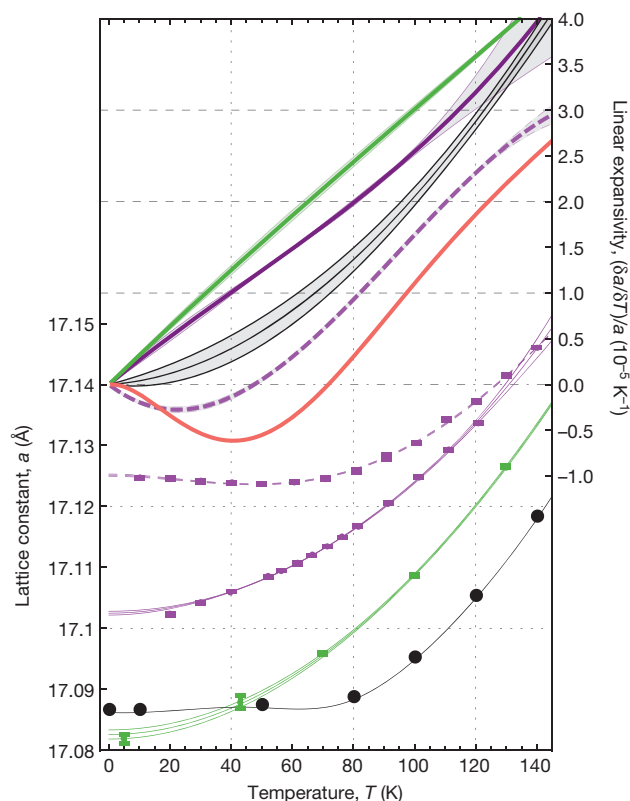**Table 1 | Structural parameters of empty and Ne-filled sII clathrate**

Atom	Position			B_{iso}	Occ.
	x	y	z		
Neon-filled sII clathrate					
O1	7/8	7/8	7/8	0.457(67)	1
O2	0.78244(10)	0.78244(10)	0.78244(10)	0.542(48)	1
O3	0.81760(6)	0.81760(6)	0.62979(11)	0.548(19)	1
D1	0.84199(20)	0.84199(20)	0.84199(20)	1.918(107)	0.5
D2	0.81604(21)	0.81604(21)	0.81604(21)	1.212(101)	0.5
D3	0.79531(16)	0.79531(16)	0.72774(21)	1.411(60)	0.5
D4	0.80466(17)	0.80466(17)	0.68471(21)	1.166(58)	0.5
D5	0.85861(12)	0.85861(12)	0.62864(27)	1.572(60)	0.5
D6	0.73076(15)	0.85494(15)	0.58776(16)	1.515(39)	0.5
Ne1	0	0	0	5.668(405)	0.86(2)
Ne2	0.42489(100)	0.42489(105)	0.42489(100)	6.459(999)	1.18(7)
Empty sII clathrate					
O1	7/8	7/8	7/8	0.621(121)	1
O2	0.78276(17)	0.78276(17)	0.78276(17)	0.590(81)	1
O3	0.81756(11)	0.81756(11)	0.62949(19)	0.584(31)	1
D1	0.84083(35)	0.84083(35)	0.84083(35)	2.507(210)	0.5
D2	0.81543(33)	0.81543(33)	0.81543(33)	1.282(173)	0.5
D3	0.79583(26)	0.79583(26)	0.72782(35)	1.514(102)	0.5
D4	0.80398(28)	0.80398(28)	0.68412(33)	1.188(97)	0.5
D5	0.85816(20)	0.85816(20)	0.62880(43)	1.585(100)	0.5
D6	0.73137(26)	0.85451(24)	0.58684(27)	1.494(65)	0.5

The crystallographic information files (CIF) of both structure determinations are given in Supplementary Information. Occupancy (Occ.) is the occupancy of a crystallographic site as compared to full occupation. The isotropic Debye–Waller factor is given as $B_{iso} = 8\pi^2 \langle u_{iso}^2 \rangle$, where $\langle u_{iso}^2 \rangle$ is the averaged atomic mean-square displacement. In parentheses are given the corresponding digits of the estimated standard deviation (e.s.d.) as resulting from the Rietveld refinement, for example, 0.78276(17) indicates 0.78276 ± 0.00017. Although more than five Ne clathrate samples (of about 500 mg each) and the resulting empty clathrate samples have been synthesized and crystallographically investigated, the e.s.d. as shown here correspond to the Rietveld refinement of one specimen, as is common practise in the presentation of powder diffraction results.

Figure 2 | Lattice constants and linear expansivity of the hydrates investigated. Left-hand y axis, lower four curves: lattice constant a versus temperature T (data points) and polynomial fits (curves) of sII hydrates as follows; empty (magenta, dashed line), Ne-filled (magenta, solid line) and N_2-filled (green) with confidence bands ($\pm 1\sigma$); the error bars of the lattice constant data points represent the estimated standard deviation of the Rietveld refinements. Black points, with a cubic spline interpolation as a guide to the eye, correspond to predicted values from lattice dynamical MD work[4]. Right-hand y axis, upper four curves: isotropic linear expansivity ($\partial a/\partial T)/a$ of sII hydrates as follows; empty (magenta, dashed), Ne-filled (magenta, solid) and N_2-filled (green) with confidence bands ($\pm 1\sigma$) from polynomial fits of lattice constants. The red curve describes the expansivity of deuterated ice Ih, the black one a polynomial fit to the predicted values[4] with confidence bands ($\pm 1\sigma$).

contrast to filled clathrates, the empty hydrate shows a marked negative thermal expansion at temperatures below ~55 K (Fig. 2) that is characteristic for open tetrahedrally-bonded framework structures[19,20] formed by water, SiO$_2$, Si or Ge. This phenomenon is due to low-energy framework phonons that tend to shorten the bond distances at low temperatures, before the anharmonicity of the remaining higher-energy lattice modes leads to a normal thermal expansion as shown for ice Ih[21–23]. In filled hydrates, the corresponding hydrogen-bond bending modes are more restricted due to the volume excluded by the guest molecules, thus preventing the framework-intrinsic negative thermal expansion.

The empty hydrate we report is the seventeenth crystalline phase of water that has been experimentally realized so far, and the one with the lowest density (0.81 g cm^{-3} for the equivalent H$_2$O form at 5 K). Following established rules concerning ice nomenclature[11,24], and considering that the designation of ice phases Ih (and Ic) up to XV is generally accepted, it should be named ice XVI. The empty hydrate has been proposed[7,12] to be one of the low-temperature water phases that is stable at negative pressures between ~0.4 and 1 GPa, where it melts at high temperature. The thermodynamically stable phase of water at the conditions of our study undoubtedly is ice Ih; but the empty clathrate is at least mechanically stable on laboratory timescales up to temperatures of ~145 K, when it starts to decompose into stacking disordered ice I[25]. This decomposition transition takes place at similar temperatures to the transformation of high-pressure phases of ice that have been recovered to ambient pressure at low temperature, with that transformation linked to the onset of orientational degrees of freedom of the water molecules[8,11]. We note that the Ne-filled sII clathrate starts to decompose at some 20 K higher (at ~165 K) than the empty clathrate, where the lack of guest molecules leaves additional operational space so that hydrogen-bond breaking rotational motion of water molecules (for example, mediated by migrating Bjerrum defects[8]) can form new water topologies.

Statistical thermodynamic theory[2] and molecular calculations[5,7] suggest that an empty hydrate with the topology of the sII clathrate will be more stable than an empty water hydrate with the topology of the sI clathrate or with any other conceivable clathrate topology[6]. This has

been ascribed[4] in part to the preponderance in the sII topology of pentagons with less strained hydrogen-bond angles (as compared to the angles in planar hexagons, which are more prominent in the sI topology), yet the situation is more complex than this[6]. The averaged hydrogen-bond distance of 2.751 Å and the averaged hydrogen-bond angle of 109.4° of the empty sII hydrate are very close to the ice Ih values of 2.750 Å and 109.5°, respectively, but their spread is distinctly larger in the sII framework (ranging from 2.738 to 2.785 Å and 105.5° to 119.8°, respectively). Filling the cages with Ne reduces the extent of water molecule displacements (by ~8%, compare with the isotropic Debye–Waller factor of oxygen (B_{iso}) in Table 1), illustrating the importance of the excluded volume or kinetic effect[26] in the filled hydrate. Like in ice Ih, the large molecular displacements arise from vibrational contributions and static disorder that arises as the more or less strained local hydrogen-bond geometries (that is, deviations from the lowest energy hydrogen-bond angles and distances) pull oxygen atoms out of their crystallographic high-symmetry time–space averaged mean positions[27]. Only a proper configurational sampling in quantum chemical calculations will establish reliable hydrogen-bond geometries[27].

A comparison of the empty and Ne-filled sII structures confirms the importance of the interaction between the water host and its Ne guest: we see a substantial volume reduction of 0.4% for the hydrogen-bonded water framework at 5 K (Fig. 2), with the magnitude of this effect similar

to lattice dynamical molecular dynamics predictions[4]. We find that the volume of the small 5^{12} cage is reduced by an even larger 0.5%, which contrasts with the identical volumes of an isolated dodecahedral 5^{12} water cluster with and without Ne found in recent *ab initio* calculations[28]. Volumetric changes arising from guest–host interactions are also not accounted for in gas hydrate prediction programs[3]; this could be corrected by using the molar volumes obtained for the empty sII hydrate to improve volume-sensitive predictions (concerning the cage fillings for pure and mixed gas hydrates, or binary hydrate phase equilibria, for example).

The original[2] and also more recent[1] statistical thermodynamic treatments of clathrates do not account for guest-induced changes to the water framework, which can limit applicability[26] and give predictions inconsistent with *ab initio* results[29]. Our results clearly document the considerable shrinkage of the water framework upon inclusion of small guests, and also highlight the importance of the excluded volume or kinetic effect in determining the different mechanical stabilities of filled and empty hydrates at atmospheric pressure (that is, outside their field of thermodynamic stability) and in causing the negative thermal expansion of the empty hydrate. It is noteworthy in this context that all water clathrate topologies are the result of a systematic hydrophobic hydration of the encaged apolar guest molecules. A structure with sI clathrate hydrate topology is the best solution to the so-called Kelvin problem[30] of minimizing the partitional area with a given amount of fully hydrogen-bonded water molecules; this in turn means that the water molecules can interact with several hydrophobic guests located in all adjacent partitions (cages). Both kinetic (vibrational) and potential energy contributions are important for hydrate stability; they should now be further quantified for a better understanding of gas hydrates and the prediction of their phase diagrams, composition and physical properties (for example, thermal conductivity). The established empty hydrate water framework provides a computational as well as an experimental reference for such future efforts.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Sloan, E. D. & Koh, C. A. *Clathrate Hydrates of Natural Gases* 3rd edn (CRC Press, Taylor & Francis Group, 2008).
2. van der Waals, J. H. & Platteeuw, J. C. Clathrate solutions. *Adv. Chem. Phys.* **2,** 1–57 (1959).
3. Ballard, A. L. & Sloan, E. D. The next generation of hydrate prediction I. Hydrate standard states and incorporation of spectroscopy. *Fluid Phase Equilib.* **194–197,** 371–383 (2002).
4. Belosludov, V. R. *et al.* Thermal expansion and lattice distortion of clathrate hydrates of cubic structures I and II. *J. Supramol. Chem.* **2,** 453–458 (2002).
5. Koyama, Y., Tanaka, H. & Koga, K. On the thermodynamic stability and structural transition of clathrate hydrates. *J. Chem. Phys.* **122,** 074503 (2005).
6. Matsumoto, M. & Tanaka, H. On the structure selectivity of clathrate hydrates. *J. Phys. Chem. B* **115,** 8257–8265 (2011).
7. Jacobson, L. C., Hujo, W. & Molinero, V. Thermodynamic stability and growth of guest-free clathrate hydrates: a low-density crystal phase of water. *J. Phys. Chem. B* **113,** 10298–10307 (2009).
8. Wooldridge, P. J., Richardson, H. H. & Devlin, J. P. Mobile Bjerrum defects — A criterion for ice-like crystal-growth. *J. Chem. Phys.* **87,** 4126–4131 (1987).
9. Guloy, A. M. *et al.* A guest-free germanium clathrate. *Nature* **443,** 320–323 (2006).
10. Gryko, J. *et al.* Low-density framework form of crystalline silicon with a wide optical band gap. *Phys. Rev. B* **62,** R7707–R7710 (2000).
11. Bartels-Rausch, T. *et al.* Ice structures, patterns, and processes: a view across the icefields. *Rev. Mod. Phys.* **84,** 885–944 (2012).
12. Conde, M. M., Vega, C., Tribello, G. A. & Slater, B. The phase diagram of water at negative pressures: Virtual ices. *J. Chem. Phys.* **131,** 034510 (2009).
13. Gies, H., Liebau, F. & Gerke, H. ''Dodecasile'' - eine neue Reihe polytyper Einschlußverbindungen von SiO_2. *Angew. Chem.* **94,** 214–215 (1982).
14. Falenty, A., Salamatin, A. N. & Kuhs, W. F. Kinetics of CO_2-hydrate formation from ice powders: data summary and modeling extended to low temperatures. *J. Phys. Chem. C* **117,** 8443–8457 (2013).
15. Alavi, S. & Ripmeester, J. A. Hydrogen-gas migration through clathrate hydrate cages. *Angew. Chem. Int. Edn* **46,** 6102–6105 (2007).
16. Senadheera, L. & Conradi, M. S. Rotation and diffusion of H_2 in hydrogen – Ice clathrate by ^{1}H NMR. *J. Phys. Chem. B* **111,** 12097–12102 (2007).
17. Dyadin, Y. A. *et al.* Clathrate formation in water-noble gas (hydrogen) systems at high pressures. *J. Struct. Chem.* **40,** 790–795 (1999).
18. Mao, W. L. *et al.* Hydrogen clusters in clathrate hydrate. *Science* **297,** 2247–2249 (2002).
19. Evans, J. S. O. Negative thermal expansion materials. *J. Chem. Soc. Dalton Trans.* 3317–3326 (1999).
20. Tang, X. L. *et al.* Thermal properties of Si-136: Theoretical and experimental study of the type-II clathrate polymorph of Si. *Phys. Rev. B* **74,** 014109 (2006).
21. Röttger, K., Endriss, A., Ihringer, J., Doyle, S. & Kuhs, W. F. Lattice constants and thermal expansion of H_2O and D_2O ice Ih between 10 and 265 K. *Acta Crystallogr. B* **50,** 644–648 (1994); Addendum. *Acta Crystallogr. B* **68,** 91 (2012).
22. Tanaka, H. Thermodynamic stability and negative thermal expansion of hexagonal and cubic ices. *J. Chem. Phys.* **108,** 4887–4893 (1998).
23. Pamuk, B. *et al.* Anomalous nuclear quantum effects in ice. *Phys. Rev. Lett.* **108,** 193003 (2012).
24. Lobban, C., Finney, J. L. & Kuhs, W. F. The structure of a new phase of ice. *Nature* **391,** 268–270 (1998).
25. Kuhs, W. F., Sippel, C., Falenty, A. & Hansen, T. C. Extent and relevance of stacking disorder in ''ice I_c''. *Proc. Natl Acad. Sci. USA* **109,** 21259–21264 (2012).
26. Rodger, P. M. Lattice relaxation in type I gas hydrates. *AIChE J.* **37,** 1511–1516 (1991).
27. Kuo, J. L., Klein, M. L. & Kuhs, W. F. The effect of proton disorder on the structure of ice-Ih: A theoretical study. *J. Chem. Phys.* **123,** 134505 (2005).
28. Kumar, P. & Sathyamurthy, N. Theoretical studies of host-guest interaction in gas hydrates. *J. Phys. Chem. A* **115,** 14276–14281 (2011).
29. Anderson, B. J., Bazant, M. Z., Tester, J. W. & Trout, B. L. Application of the cell potential method to predict phase equilibria of multicomponent gas hydrate systems. *J. Phys. Chem. B* **109,** 8153–8163 (2005).
30. Weaire, D. & Phelan, R. A counterexample to Kelvin conjecture on minimal-surfaces. *Phil. Mag. Lett.* **69,** 107–110 (1994).

METHODS

Sample preparation. Ne clathrates were prepared from deuterated ice spheres with a mean diameter of $\sim 18\,\mu m$ produced by shock freezing D_2O water mist in liquid N_2. Fine water droplets were created with a commercial airbrush using 0.6 MPa of N_2 as a carrier gas. A risk of contamination with H_2O has been minimized by spraying under a protective N_2 atmosphere. Ice spheres were transferred into Al vials ($\sim 1\,cm^3$ each), inserted in a pressure cell and transformed within ~ 30 min into pure Ne clathrates by a stepwise pressurization with Ne to a final pressure of ~ 0.35 GPa at 244 K, that is, passing through the nominal ice Ih–ice II phase transition. All samples were recovered, mixed together and crushed in a stainless steel mortar at liquid N_2 temperature to increase the surface area for the leaching experiments. Final products were filled back into Al vials and stored at liquid nitrogen temperatures. Cryo-SEM images taken on Ne-hydrate powders showed no substantial shape alterations to the initial spherical ice spheres.

Neutron diffraction experiments. These were carried out on the high-intensity two-axis diffractometer D20[31] at the high-flux reactor of ILL, Grenoble, France. An Al vial with $\sim 1\,cm^3$ of deuterated Ne clathrate was inserted at liquid N_2 temperatures into a vanadium cell, connected to a vacuum pump and inserted into a temperature-controlled He-flow cryostat; temperatures were controlled within ~ 0.1 K of the desired value. The *in situ* leaching experiments were performed to study the T-dependency of the progressive emptying of the small and large clathrate cages by powder diffraction measurements at a wavelength of $\lambda \approx 2.419$ Å. The data sets were used for a Rietveld refinement[32] of the structural parameters, and delivered (amongst others) the occupancy of the small and large cages (Extended Data Fig. 1). The empty hydrate sample obtained after five days of pumping was used for a detailed structural investigation at $\lambda \approx 1.1226$ Å (Extended Data Fig. 2); a similar investigation was made for the initial Ne clathrate (Extended Data Fig. 2). Subsequently, both, empty and Ne-filled sII clathrate were subjected to a temperature dependent study of the lattice constants; similarly, a sample of sII N_2 clathrate was also studied. The temperature was ramped up in steps of typically 10 K (30 K for N_2 clathrate) between 10 and 140 K. Rietveld analysis of the diffraction patterns obtained at $\lambda \approx 2.4157$ Å revealed lattice constants with a precision of 0.01‰ (Fig. 2); the T-dependency was described by a polynomial expression fitted to the data with coefficients given in Extended Data Table 1. The expansivities were deduced from these expressions by differentiation, and are plotted in Fig. 2 together with confidence bands deduced by error propagation considering also parameter correlation. These confidence bands around the mean predictions were calculated using a Monte-Carlo approach realized by Kuhs[33], employing the 'square root' method described, for example, by James[34]. It considers the full covariance matrix of the linear model fit of the lattice parameters; this is important as the correlations between polynomial coefficients are very high. For each point of the curve, a normal distribution of 100 predicted points, considering all correlations, is generated, from which a standard deviation can be obtained straightforwardly. The mean value (which corresponds to the derived polynomial fit of the lattice parameters) $\pm$ these standard deviations give the plotted confidence bands.

Modelling of out-diffusion. The emptying of the initially Ne-filled cages. The Ne-filled clathrate consisting of a log-normally distributed assembly of micrometre-sized spheres can be modelled as a leaching process in a shrinking-core approach (Extended Data Fig. 1). The T-dependent leaching process follows Arrhenius' equation, delivering activation energies of 30.0 and 12.8 kJ mol^{-1} for small and large cages, respectively. This suggests that out-diffusion through the water hexagons of the large cages can proceed with hosts the size of Ne (Fig. 1), while a jump through the pentagonal faces of the small cages is only possible when a water vacancy is present in the ring. There is accumulated evidence from molecular dynamics simulations as well as from experimental work[14,35,36] that the bulk diffusion of larger guest molecules, like CH_4 or CO_2, proceeds via these water vacancies in a process called the 'hole-in-the-cage' mechanism. The low activation energy of the out-diffusion from the large cages (proceeding through 6-membered water rings) suggests that it involves thermally activated jumps through the intact cage wall[37].

31. Hansen, T. C., Henry, P. F., Fischer, H. E., Torregrossa, J. & Convert, P. The D20 instrument at the ILL: a versatile high-intensity two-axis neutron diffractometer. *Meas. Sci. Technol.* **19,** 034001 http://dx.doi.org/10.1088/0957-0233/19/3/034001 (2008).

32. Rodríguez-Carvajal, J. Recent advances in magnetic-structure determination by neutron powder diffraction. *Physica B* **192,** 55–69 (1993).

33. Kuhs, W. F. Generalized atomic displacements in crystallographic structure-analysis. *Acta Crystallogr. A* **48,** 80–98 (1992).

34. James, F. Monte-Carlo theory and practice. *Rep. Prog. Phys.* **43,** 1145–1189 (1980).

35. Peters, B., Zimmermann, N. E. R., Beckham, G. T., Tester, J. W. & Trout, B. L. Path sampling calculation of methane diffusivity in natural gas hydrates from a water-vacancy assisted mechanism. *J. Am. Chem. Soc.* **130,** 17342–17350 (2008).

36. Buch, V. *et al.* Clathrate hydrates with hydrogen-bonding guests. *Phys. Chem. Chem. Phys.* **11,** 10245–10265 (2009).

37. Demurov, A., Radhakrishnan, R. & Trout, B. L. Computations of diffusivities in ice and CO_2 clathrate hydrates via molecular dynamics and Monte Carlo simulations. *J. Chem. Phys.* **116,** 702–709 (2002).
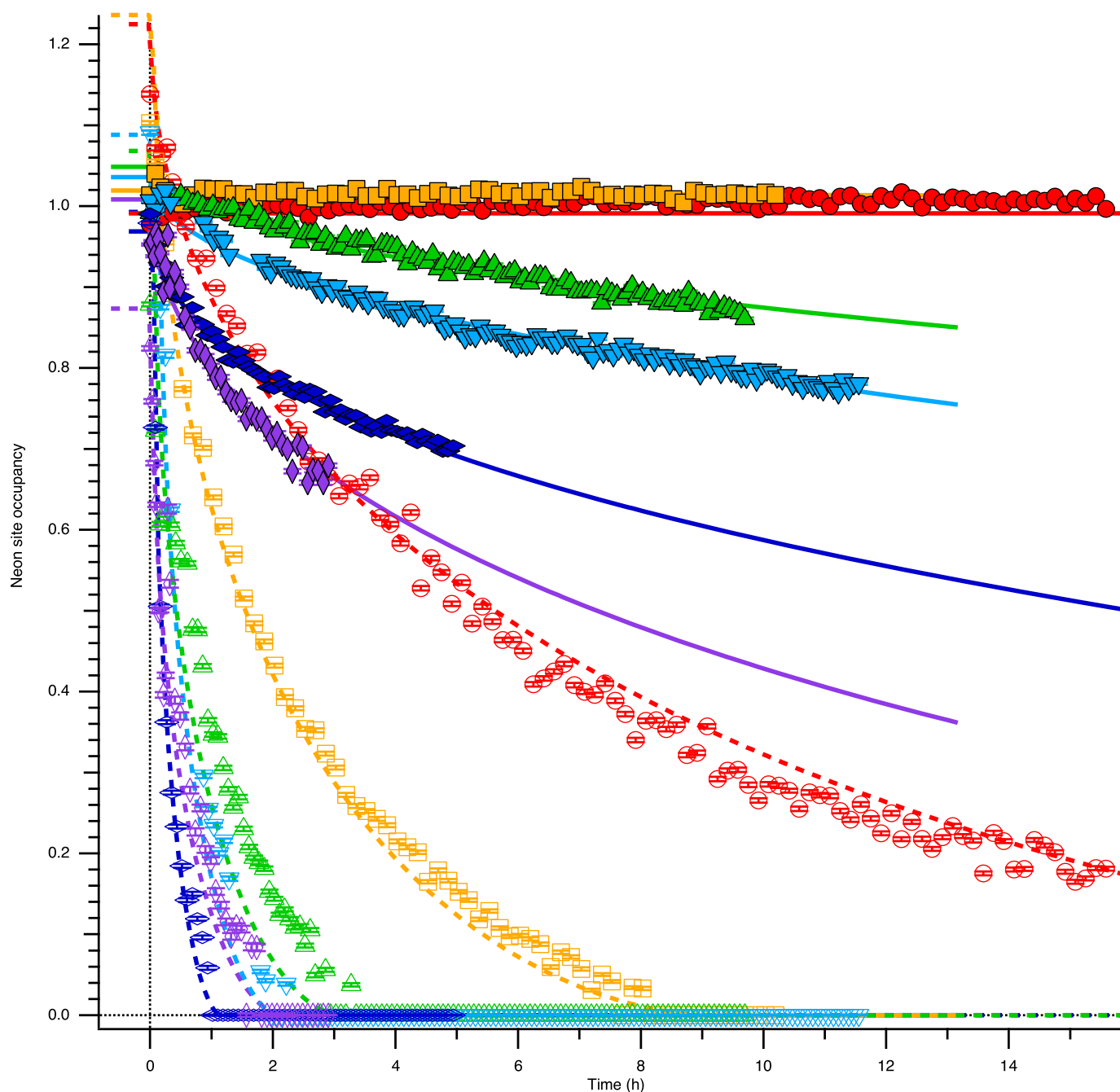
Extended Data Figure 1 | Cage filling as a function of time for different temperatures. A shrinking core model of the Ne leaching process fits values from sequential Rietveld refinement. Filled data points (and solid line for the fits) present the data for small cages (Ne1); open data points (and dashed lines) show data for the large cages (Ne2). Red circles and lines represent 110 K, orange squares and lines 120 K, green triangles and lines 130 K, light blue triangles and lines 135 K, dark blue flat rhombi and lines 140 K, magenta upright rhombi and lines 145 K.
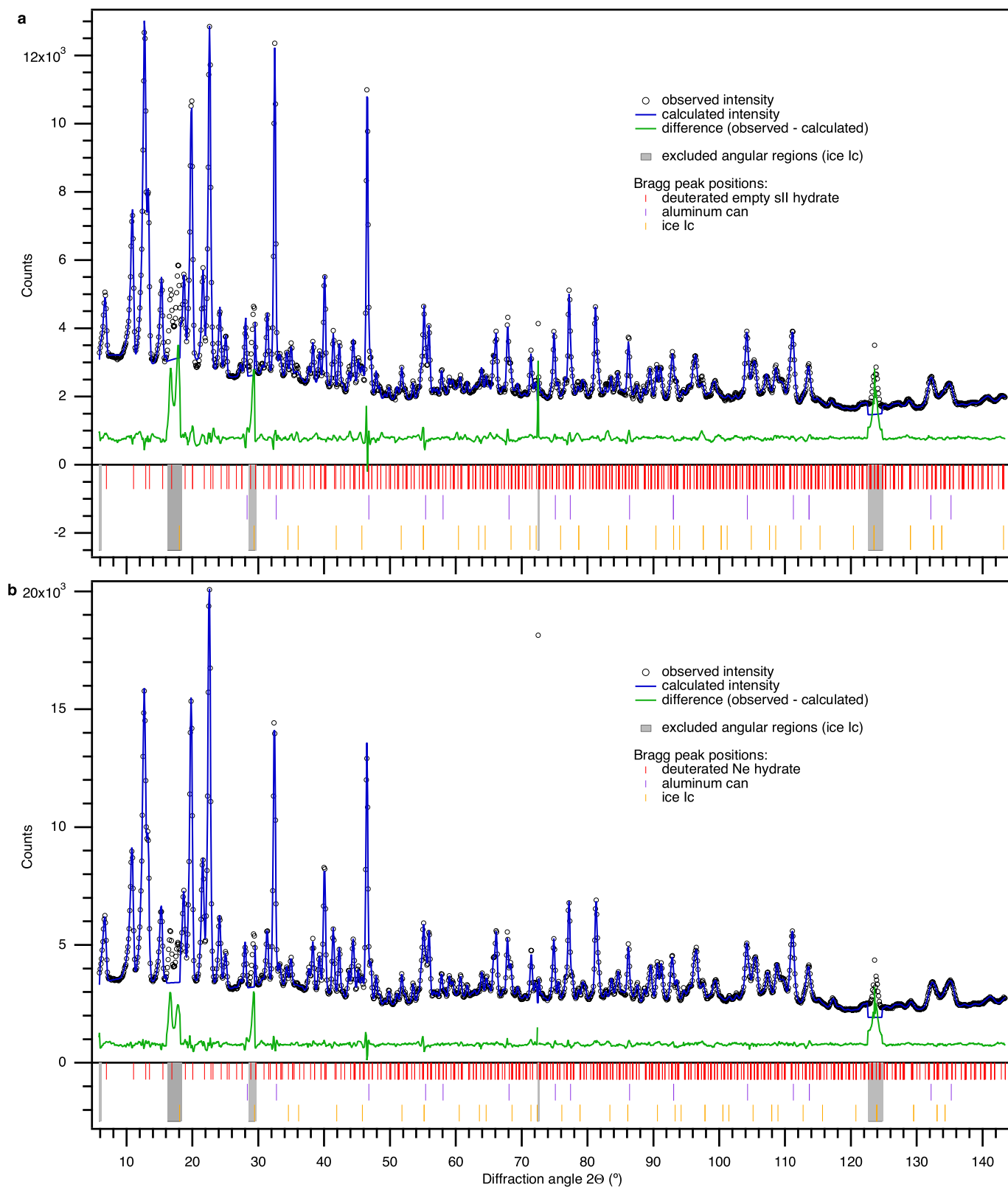
Extended Data Figure 2 | Diffraction patterns of Ne-filled and empty hydrate. a, b, Rietveld fit (obtained using FullProf software[32]) to diffraction pattern of empty sII D$_2$O hydrate (**a**) and Ne D$_2$O hydrate (**b**) taken at 5 K ($\lambda \approx 1.1226$ Å) on D20, ILL/Grenoble. The observed intensity is represented by open black circles, the calculated intensity as a blue line, the difference of both by a green line, grey shading marks the angular regions excluded in the refinement, red lines mark the positions of Bragg peaks of the hydrate, violet lines those of the aluminium sample can and orange lines those of ice Ic.

Extended Data Table 1 | Polynomial coefficients of lattice constant fits*

	Ne hydrate	empty hydrate	N$_2$ hydrate
A_0 / Å	17.102504(297)	17.125133(167)	17.082609(724)
A_1 / (Å K^{-1})	0	0	0
A_2 / (Å K^{-2})	2.328(458)·10^{-6}	-2.207(198)·10^{-6}	2.760(170)·10^{-6}
A_3 / (Å K^{-3})	-4.43(871)·10^{-9}	37.18(318)·10^{-9}	-1.83(122)·10^{-9}
A_4 / (Å K^{-4})	25.8(432)·10^{-12}	-98.4(132)·10^{-12}	4.39(236) ·10^{-12}

Parameter estimates with standard errors in brackets of the polynomial fit to the T-dependent lattice constants of the deuterated hydrates: $a(T) = A_0 + A_1 T + A_2 T^2 + A_3 T^3 + A_4 T^4$. The term A_1 is set to zero to address the fact that the linear expansivity, $\partial a/(a \cdot \partial T)$, needs to be zero at 0 K.
* From Fig. 2.

Isotopic constraints on marine and terrestrial N$_2$O emissions during the last deglaciation

Adrian Schilt[1,2], Edward J. Brook[1], Thomas K. Bauska[1], Daniel Baggenstos[3], Hubertus Fischer[2], Fortunat Joos[2], Vasilii V. Petrenko[4], Hinrich Schaefer[5], Jochen Schmitt[2], Jeffrey P. Severinghaus[3], Renato Spahni[2] & Thomas F. Stocker[2]

Nitrous oxide (N$_2$O) is an important greenhouse gas and ozone-depleting substance that has anthropogenic as well as natural marine and terrestrial sources[1]. The tropospheric N$_2$O concentrations have varied substantially in the past in concert with changing climate on glacial–interglacial and millennial timescales[2–8]. It is not well understood, however, how N$_2$O emissions from marine and terrestrial sources change in response to varying environmental conditions. The distinct isotopic compositions of marine and terrestrial N$_2$O sources can help disentangle the relative changes in marine and terrestrial N$_2$O emissions during past climate variations[4,9,10]. Here we present N$_2$O concentration and isotopic data for the last deglaciation, from 16,000 to 10,000 years before present, retrieved from air bubbles trapped in polar ice at Taylor Glacier, Antarctica. With the help of our data and a box model of the N$_2$O cycle, we find a 30 per cent increase in total N$_2$O emissions from the late glacial to the interglacial, with terrestrial and marine emissions contributing equally to the overall increase and generally evolving in parallel over the last deglaciation, even though there is no a priori connection between the drivers of the two sources. However, we find that terrestrial emissions dominated on centennial timescales, consistent with a state-of-the-art dynamic global vegetation and land surface process model that suggests that during the last deglaciation emission changes were strongly influenced by temperature and precipitation patterns over land surfaces. The results improve our understanding of the drivers of natural N$_2$O emissions and are consistent with the idea that natural N$_2$O emissions will probably increase in response to anthropogenic warming[11].

Ice-core studies indicate that during the past 800 kyr tropospheric N$_2$O concentrations ranged from about 200 to 300 p.p.b., covarying with climate on glacial–interglacial and millennial timescales (refs 2–8 and Fig. 1a). Pre-industrial atmospheric N$_2$O concentrations were regulated by microbiological production in marine and terrestrial environments and by photochemical destruction in the stratosphere[1]. Simulations suggest that the pre-industrial atmospheric lifetime of 142 ± 14 yr (ref. 12) remained relatively constant over the last deglaciation[13,14], and, therefore, past atmospheric N$_2$O concentrations were mainly modulated by emission strength. Importantly, emission strength increases in a warmer climate, implying that natural ecosystem N$_2$O production constitutes a positive climate feedback that will add to the anthropogenic N$_2$O load in the atmosphere and amplify the warming in coming centuries. However, the details of the relative sensitivities of marine and terrestrial sources to changing environmental conditions are not known at present, hampering quantitative projections of future emissions.
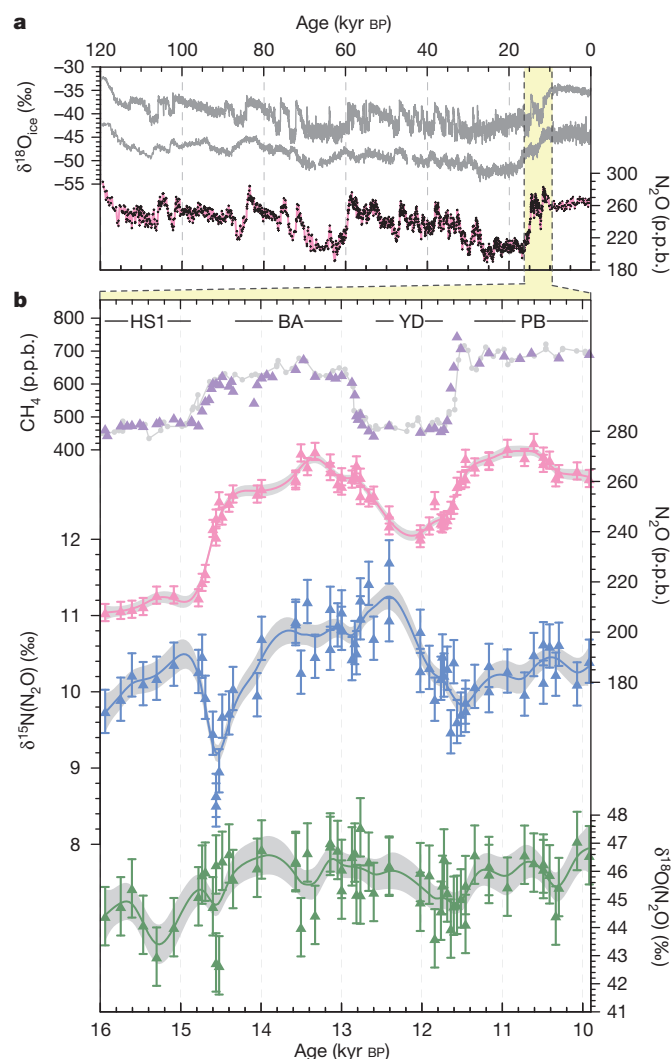
Modern field data indicate that marine and terrestrial N$_2$O emissions exhibit distinct isotopic compositions (Extended Data Fig. 1), with marine N$_2$O being more enriched in both heavy isotopes (δ^{15}N of 4–12‰ relative to atmospheric N$_2$; δ^{18}O of 42–67‰ relative to VSMOW) than is terrestrial N$_2$O (δ^{15}N of −34–2‰; δ^{18}O of 20–43‰). Therefore, the isotopic composition of tropospheric N$_2$O is a powerful tool for disentangling the relative changes of marine and terrestrial N$_2$O emissions

during past climate variations[4,9,10]. For instance, decreasing tropospheric δ^{15}N and δ^{18}O indicate increasing importance of terrestrial emissions, whereas increasing δ^{15}N and δ^{18}O indicate increasing importance of marine emissions. The only prior study of the N$_2$O isotopic composition over the last deglaciation[4] found minimal change in the ratio of marine to terrestrial N$_2$O emissions, but was hindered by the relatively low temporal sampling resolution and precision of the measurements.

We determined the concentration and the isotopic composition of N$_2$O over the last deglaciation from 16–10 kyr before present (BP, AD 1950) (Fig. 1b) from a total of 64 ice samples collected along a horizontal transect on the Taylor Glacier, Antarctica (D.B. *et al.*, manuscript in preparation). Measurement precision, as determined by replicate analyses and reported here as pooled standard deviations (1σ), was 3.4 p.p.b. for the N$_2$O concentration, 0.28‰ for δ^{15}N and 1.04‰ for δ^{18}O. The mean temporal sampling resolution was better than 100 yr. The timescale was established by synchronizing the fast global changes in tropospheric methane (CH$_4$) concentrations in the Taylor Glacier data with the corresponding data of the WAIS Divide deep ice core on an updated version of the WDC06A-7 timescale[15], with further constraints from the isotopic composition of atmospheric molecular oxygen between fast CH$_4$ changes (Methods). Note that atmospheric data extracted from polar ice samples are a smoothed representation of the atmospheric history owing to the mixing of air in the firn column; Taylor Glacier is expected to have a gas age distribution with a range of about 300 yr, similar to the Taylor Dome ice core[16]. The new Taylor Glacier N$_2$O isotopic data (δ^{15}N and δ^{18}O) measured at Oregon State University agree with the only previously published N$_2$O isotopic data covering the last deglaciation, from the Taylor Dome ice core[4], as well as with new measurements performed at the University of Bern on ice samples from the Taylor Glacier and the Talos Dome ice core (Extended Data Fig. 2 and Methods). Taylor Glacier measurements confirm the trends in tropospheric N$_2$O concentrations from previous ice-core studies[2–5], and show the following general features in high temporal sampling resolution (Fig. 1b): N$_2$O rapidly increased from 211 ± 1 p.p.b. (mean ± s.e. in the time interval 15.9–14.9 kyr BP) during Heinrich stadial 1 (HS1) to 263 ± 2 p.p.b. (14.3–13.0 kyr BP) during the Bølling–Allerød interstadial. Following a decrease to 243 ± 2 p.p.b. (12.6–11.7 kyr BP) during the Younger Dryas stadial, N$_2$O reached 267 ± 1 p.p.b. (11.3–9.9 kyr BP) during the Preboreal stage. δ^{15}N averaged 10.3 ± 0.1‰ over the last deglaciation, with excursions of up to about 2‰. Approximately similar values were reached during HS1, the Younger Dryas and the Preboreal (10.3 ± 0.1‰), but δ^{15}N was higher during the Bølling–Allerød (10.7 ± 0.1‰). δ^{18}O averaged 45.5 ± 0.1‰ over the last deglaciation, with the magnitude of variability roughly similar to the precision of the measurements.

The robust isotopic variations in the new Taylor Glacier data reveal how the major environmental changes during the last deglaciation perturbed the nitrogen cycle and N$_2$O production. Broadly, the N$_2$O concentration data imply an increase in total N$_2$O emissions of about 30% from the late glacial to the interglacial, and the similar isotopic values

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. [2]Climate and Environmental Physics, Physics Institute, and Oeschger Centre for Climate Change Research, University of Bern, 3012 Bern, Switzerland. [3]Scripps Institution of Oceanography, University of California, San Diego, California 92037, USA. [4]Department of Earth and Environmental Sciences, University of Rochester, Rochester, New York 14627, USA. [5]National Institute of Water and Atmospheric Research, Wellington 6021, New Zealand.

fast increase in terrestrial emissions that preceded the increase in marine emissions.

To estimate the evolution of total N_2O emissions as well as the relative contributions of marine and terrestrial ecosystems over the last deglaciation on the basis of the Taylor Glacier data, we used a two-box model of atmospheric N_2O and its isotopic composition (Fig. 2). The model included a well-mixed troposphere and stratosphere, separate marine and terrestrial N_2O sources, and a stratospheric sink (Methods). Investigations with a more complex formulation of the box model including an explicit representation of the marine N_2O cycle and inventory suggest that the physical effects of air–sea interactions and ocean mixing on N_2O can be neglected on the timescales of the last deglaciation (Extended Data Fig. 3). We derived plausible emission histories using only the N_2O concentration and $\delta^{15}N$ data (which offered a higher signal-to-noise ratio than the $\delta^{18}O$ data); however, the results were consistent *post hoc* with the $\delta^{18}O$ constraints (Extended Data Fig. 4). The large range of the isotopic values for both marine and terrestrial N_2O emissions observed in modern field data (Extended Data Fig. 1) precluded a quantification of the exact marine and terrestrial fractions of the total emissions based on the measured tropospheric isotopic values. Notably, the reported late pre-industrial (AD 1750) $\delta^{15}N$ value[17] is similar to the Taylor Glacier value at 16 kyr BP (Fig. 2), suggesting similar relative strengths of marine and terrestrial N_2O emissions for pre-industrial and late glacial climate conditions. We therefore prescribed an initial marine fraction of 37% (that is, an initial terrestrial fraction of 63%) of the total emissions at 16 kyr BP, in line with best estimates for the modern natural N_2O budget[1]. The sensitivity of our results to the chosen relative strength of marine and terrestrial emissions at 16 kyr BP is illustrated by further scenarios with low and high estimates of the initial marine fraction of 17% and 74% (ref. 1), demonstrating that for all scenarios the marine and terrestrial fractions showed similar trends with absolute changes of only 7% or less over the last deglaciation (Extended Data Fig. 5). The box model accounts for atmospheric imbalances (non-steady-state conditions) affecting the tropospheric concentrations and isotopic compositions at times of rapidly changing atmospheric N_2O load (Methods). However, only a small part of the observed changes in N_2O and $\delta^{15}N$ were caused by such atmospheric imbalances, indicating that changes in marine and terrestrial emissions were mostly responsible for the observed variability in the Taylor Glacier data (Fig. 2). Our approach assumes that the isotopic compositions of marine and terrestrial N_2O emissions remained constant over time. For the marine source, support for this assumption comes from the observation that the global mean isotopic composition of bioavailable nitrogen did not change significantly over the last deglaciation[18]. For the terrestrial source, a global compilation of lacustrine $\delta^{15}N$ sedimentary data[19] does not reveal substantial changes, for example in response to the Younger Dryas. However, a long-term decrease with a rate of 0.25‰ per millennium is observed from 15 to 7 kyr BP; if transferred directly to the isotopic composition of the terrestrial N_2O source, this decrease would require an increase in marine emissions at the expense of terrestrial emissions of about 0.05 Tg N yr^{-1} per millennium to keep the isotopic composition of the troposphere constant. Although not negligible, this would still be small compared with the inferred changes in total N_2O emissions, which are of the order of 2.3 Tg N yr^{-1}.

Today, the strongest marine N_2O emissions occur in the eastern tropical and northern Pacific Ocean, the Southern Ocean, the Arabian Sea, and in coastal and equatorial upwelling regions, as inferred from inverse modelling[20]. High N_2O production rates in these regions and in the global ocean as a whole are closely linked to hypoxia (low concentrations of dissolved oxygen), which is controlled by the temperature- and salinity-dependent oxygen solubility, the cycling of organic matter, the availability of nutrients and, thus, ocean circulation. Indeed, ocean models suggest that a weakening of the Atlantic meridional overturning circulation (AMOC) leads to a decrease in marine N_2O production almost everywhere in the global oceans, in particular in the low-oxygen regions, because of higher stratification, increased oceanic storage of

Figure 1 | Changes in tropospheric N_2O and climate proxies during the last glacial–interglacial cycle and the last deglaciation. a, The past 120 kyr on the AICC2012 timescale[26]: temperature proxies $\delta^{18}O_{ice}$ ($\delta^{18}O_{ice} = ((^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{VSMOW} - 1) \times 1,000‰$; VSMOW, Vienna Standard Mean Ocean Water) of Greenland (upper grey curve; North Greenland Ice Core Project[27]) and Antarctica (lower grey curve; EPICA Dronning Maud Land[28]), as well as tropospheric N_2O (pink; EPICA Dome C[29] and North Greenland Ice Core Project[3,5,8]). **b,** Detailed data from the last deglaciation from 16 to 10 kyr BP on an updated version of the WDC06A-7 timescale[15] (Methods): Taylor Glacier CH_4 (purple triangles) together with Talos Dome CH_4 (grey circles[30]), Taylor Glacier N_2O (pink), as well as $\delta^{15}N$ ($\delta^{15}N = ((^{15}N/^{14}N)_{sample}/(^{15}N/^{14}N)_{atmospheric N_2} - 1) \times 1,000‰$) (blue) and $\delta^{18}O$ (green; relative to VSMOW) of N_2O. Solid lines show splines with a cut-off period of 600 yr through the N_2O concentration and isotopic composition data. Error bars indicate pooled standard deviations of replicates ($\pm 1\sigma$, $n = 10$); grey shaded areas indicate $\pm 1\sigma$ envelopes from the Monte Carlo approach (Methods). BA, Bølling–Allerød; YD, Younger Dryas; PB, Preboreal.

during HS1 and the Preboreal indicate that both marine and terrestrial emissions contributed about equally to the overall increase. However, variations in the isotopic composition related to climate oscillations on millennial and centennial timescales during the last deglaciation point to substantially asynchronous responses of marine and terrestrial emissions on shorter timescales. The variations can be attributed to various drivers, such as changes in oxygen inventories and circulation in the global oceans, as well as changes in temperature and precipitation patterns over land. For instance, higher $\delta^{15}N$ during most of the Bølling–Allerød indicates relatively strong marine emissions, whereas the short-term decrease in $\delta^{15}N$ at the beginning of the Bølling–Allerød points to a

Figure 2 | N₂O emissions during the last deglaciation. a, Total N₂O emissions. **b,** Marine (blue) and terrestrial (green) N₂O emission changes relative to 16 kyr BP. Total, marine and terrestrial emissions were inversely calculated using the box model such that they recover the Taylor Glacier N₂O and δ¹⁵N splines (solid lines in **c** and **d**, respectively) in a forward calculation. The uncertainty bands related to the emissions result from the Monte Carlo approach and indicate ±1σ of all solutions. The absolute changes in marine and terrestrial emissions depend on the initial marine fraction, which was set to 37% of the total emissions at 16 kyr BP (see Extended Data Fig. 5 for sensitivity studies). **c,** Taylor Glacier N₂O, with ±1σ error bars. **d,** Taylor Glacier δ¹⁵N of N₂O, with ±1σ error bars. The orange dashed line indicates pre-industrial (AD 1750) δ¹⁵N (ref. 17). The dashed pink (**c**) and blue (**d**) lines show N₂O and δ¹⁵N calculated using the modelled marine and terrestrial emissions but assuming equilibrium with respect to the sink at any time (Methods); the differences between solid and dashed lines indicate the effect of atmospheric imbalances. **e,** AMOC changes estimated using the Bern3D Earth System Model (including the ±1σ uncertainty band), constrained by proxy data[23]. **f,** Terrestrial N₂O emission changes independently inferred from LPX-Bern. **g,** TraCE-21ka temperature changes over land surfaces[25] used to force LPX-Bern.

N₂O, less upwelling of nutrients into the euphotic zone, decreased primary productivity and increased subsurface oxygen concentrations impeding N₂O production[21,22].

The changes in marine N₂O emissions inferred from the Taylor Glacier data thus reflect important aspects of the globally integrated physical and biogeochemical ocean response to changing climate conditions. Reconstructions of AMOC changes[23] and the qualitative evolution of the marine oxygen inventory from a compilation of globally distributed marine sediment cores[24] can be combined with our data to provide a consistent history of marine N₂O emissions coupled to oxygen concentrations in the upper ocean. During the transition from HS1 to the Bølling–Allerød, the marine sediment data indicate a large expansion of hypoxia almost everywhere in the upper ocean, including the northern parts of the Pacific and Indian oceans, which are important regions for marine N₂O emissions[20]. Our reconstructions show that marine emissions substantially contributed to the concentration increase from 211 to 263 p.p.b. during that time (Fig. 2 and Extended Data Fig. 5); however, terrestrial emissions were similarly important (see below). During the Younger Dryas, the decrease in marine N₂O emissions occurred in concert with the weakening of the AMOC[23] as expected from model simulations[21,22]. N₂O concentrations reached slightly higher values during the Preboreal than during the Bølling–Allerød, whereas marine N₂O emissions were probably highest during the Bølling–Allerød, as also reflected by the oxygen availability in the global oceans reaching its lowest value during that time[24]. In contrast, terrestrial N₂O emissions were probably stronger during the Preboreal than during the Bølling–Allerød (Fig. 2), which may indicate that climate conditions on land increasingly favoured terrestrial N₂O emissions throughout the last deglaciation (interrupted by the Younger Dryas). This is indirectly supported by the observation that CH₄, which is controlled primarily by temperature- and precipitation-driven terrestrial sources, was also higher during the Preboreal than during the Bølling–Allerød (Fig. 1b). Although the general trends in marine and terrestrial N₂O emissions were coupled over the last deglaciation, confirming that both sources substantially contributed to the observed concentration increase[4], there are important differences on shorter timescales (Fig. 2). Notably, at the beginning of the Bølling–Allerød, the strong decrease in δ¹⁵N suggests that terrestrial emissions increased more rapidly than did marine emissions, and reached an early maximum between 15 and 14 kyr BP.

For an independent comparison with the reconstructed emissions inferred from the Taylor Glacier data, we simulated terrestrial N₂O emissions using LPX-Bern, a dynamic global vegetation and land surface process model[11] (Methods). LPX-Bern was forced with climate anomalies from the TraCE-21ka experiment, a general circulation model simulation of climate over the last deglaciation with orbital, greenhouse gas, ice-sheet and meltwater forcings[25]. The LPX-Bern simulations qualitatively reproduce the reconstructed trends as well as variations on millennial and centennial timescales in terrestrial N₂O emissions

over the last deglaciation, including the early maximum between 15 and 14 kyr BP, the decrease from the Bølling–Allerød into the Younger Dryas, and the subsequent increase into the Preboreal to values slightly above those of the Bølling–Allerød (Fig. 2). The simulated emission changes in LPX-Bern are strongly influenced by temperature (and precipitation) patterns over land surfaces, suggesting that those parameters contributed to the variability in terrestrial N_2O emissions over the last deglaciation. The qualitative agreement between the reconstructed and modelled emissions, coupled with the apparent sensitivity of N_2O emissions to temperature in both the model simulation and ice-core data, strongly suggests that terrestrial emissions acted as a positive feedback on climate change during the last deglaciation.

Our results provide insight into the nitrogen cycle and the overall functioning of marine and terrestrial ecosystems under varying environmental conditions, and are consistent with the hypothesis that natural N_2O emissions will probably increase in response to anthropogenic warming. The results also increase confidence in the ability of present-generation dynamic global vegetation and land surface process models to project changes in terrestrial N_2O emissions in response to climate change.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Stocker, T. F. et al. (eds) Climate Change 2013: The Physical Science Basis (Cambridge Univ. Press, 2013).
2. Flückiger, J. et al. Variations in atmospheric N_2O concentration during abrupt climatic changes. Science **285**, 227–230 (1999).
3. Schilt, A. et al. The response of atmospheric nitrous oxide to climate variations during the last glacial period. Geophys. Res. Lett. **40**, 1888–1893 (2013).
4. Sowers, T., Alley, R. B. & Jubenville, J. Ice core records of atmospheric N_2O covering the last 106,000 years. Science **301**, 945–948 (2003).
5. Schilt, A. et al. Atmospheric nitrous oxide during the last 140,000 years. Earth Planet. Sci. Lett. **300**, 33–43 (2010).
6. Spahni, R. et al. Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores. Science **310**, 1317–1321 (2005).
7. Schilt, A. et al. Glacial-interglacial and millennial-scale variations in the atmospheric nitrous oxide concentration during the last 800,000 years. Quat. Sci. Rev. **29**, 182–192 (2010).
8. Flückiger, J. et al. N_2O and CH_4 variations during the last glacial epoch: insight into global processes. Glob. Biogeochem. Cycles **18**, GB1020 (2004).
9. Ishijima, K. et al. Temporal variations of the atmospheric nitrous oxide concentration and its $\delta^{15}N$ and $\delta^{18}O$ for the latter half of the 20th century reconstructed from firn air analyses. J. Geophys. Res. **112**, D03305 (2007).
10. Röckmann, T., Kaiser, J. & Brenninkmeijer, C. A. M. The isotopic fingerprint of the pre-industrial and the anthropogenic N_2O source. Atmos. Chem. Phys. **3**, 315–323 (2003).
11. Stocker, B. D. et al. Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. Nature Clim. Change **3**, 666–672 (2013).
12. Prather, M. J., Holmes, C. D. & Hsu, J. Reactive greenhouse gas scenarios: systematic exploration of uncertainties and the role of atmospheric chemistry. Geophys. Res. Lett. **39**, L09803 (2012).
13. Crutzen, P. J. & Bruhl, C. A model study of atmospheric temperatures and the concentrations of ozone, hydroxyl, and some other photochemically active gases during the glacial, the preindustrial Holocene and the present. Geophys. Res. Lett. **20**, 1047–1050 (1993).
14. Martinerie, P., Brasseur, G. P. & Granier, C. The chemical composition of ancient atmospheres: a model study constrained by ice core data. J. Geophys. Res. **100**, 14291–14304 (1995).
15. Wais Divide Project Members. Onset of deglacial warming in West Antarctica driven by local orbital forcing. Nature **500**, 440–444 (2013).
16. Indermühle, A. et al. Atmospheric CO_2 concentration from 60 to 20 kyr BP from the Taylor Dome Ice Core, Antarctica. Geophys. Res. Lett. **27**, 735–738 (2000).
17. Bernard, S. et al. Constraints on N_2O budget changes since pre-industrial time from new firn air and ice core isotope measurements. Atmos. Chem. Phys. **6**, 493–503 (2006).
18. Galbraith, E. D., Kienast, M. & The NICOPP working group members. The acceleration of oceanic denitrification during deglacial warming. Nature Geosci. **6**, 579–584 (2013).
19. McLauchlan, K. K., Williams, J. J., Craine, J. M. & Jeffers, E. S. Changes in global nitrogen cycling during the Holocene epoch. Nature **495**, 352–355 (2013).
20. Hirsch, A. I. et al. Inverse modeling estimates of the global nitrous oxide surface flux from 1998–2001. Glob. Biogeochem. Cycles **20**, GB1008 (2006).
21. Goldstein, B., Joos, F. & Stocker, T. F. A modeling study of oceanic nitrous oxide during the Younger Dryas cold period. Geophys. Res. Lett. **30**, 1092 (2003).
22. Schmittner, A. & Galbraith, E. D. Glacial greenhouse-gas fluctuations controlled by ocean circulation changes. Nature **456**, 373–376 (2008).
23. Ritz, S. P. et al. Estimated strength of the Atlantic overturning circulation during the last deglaciation. Nature Geosci. **6**, 208–212 (2013).
24. Jaccard, S. L. & Galbraith, E. D. Large climate-driven changes of oceanic oxygen concentrations during the last deglaciation. Nature Geosci. **5**, 151–156 (2012).
25. Liu, Z. et al. Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. Science **325**, 310–314 (2009).
26. Veres, D. et al. The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. Clim. Past **9**, 1733–1748 (2013).
27. NGRIP Members. High-resolution record of Northern Hemisphere climate extending into the last interglacial period. Nature **431**, 147–151 (2004).
28. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. Nature **444**, 195–198 (2006).
29. Flückiger, J. et al. High-resolution Holocene N_2O ice core record and its relationship with CH_4 and CO_2. Glob. Biogeochem. Cycles **16**, 1010 (2002).
30. Stenni, B. et al. Expression of the bipolar see-saw in Antarctic climate records during the last deglaciation. Nature Geosci. **4**, 46–49 (2011).

Author Contributions J.P.S., V.V.P. and E.J.B. initiated and led the Taylor Glacier project. T.K.B., E.J.B., D.B., V.V.P., H.S., A.S. and J.P.S. performed fieldwork. A.S. set up the experimental apparatus for N_2O isotopes and carried out the measurements, with great support from T.K.B., E.J.B. and J.S. Intercomparison measurements were performed by J.S. at the University of Bern and by T.K.B. at Oregon State University. D.B. and T.K.B. developed the timescale for Taylor Glacier. A.S. and F.J. performed the box modelling. R.S. and F.J. produced and interpreted the LPX-Bern model results. All authors discussed the results and contributed to the manuscript, which was written by A.S.

METHODS

Standard gases for N_2O concentration and isotopic composition. To date, no official international standard gases exist for the isotopic composition of N_2O, and isotopic results are instead reported directly on the internationally accepted atmospheric N_2 (for $\delta^{15}N$) and VSMOW (for $\delta^{18}O$) scales. For practical reasons it is advantageous to have a standard gas consisting of N_2O in air at tropospheric concentration. Therefore, a standard gas cylinder of tropospheric background air, labelled NOAA-1, filled on 11 December 2008 at Niwot Ridge, Colorado, was used as the primary standard at Oregon State University. The N_2O concentration was 322.32 ± 0.14 p.p.b. according to a calibration by the National Oceanic and Atmospheric Administration (NOAA-2006A scale). The isotopic composition was expressed in the customary delta notation, that is, $\delta^{15}N = ({}^{15}R_{sample}/{}^{15}R_{standard} - 1) \times 1{,}000‰$ and $\delta^{18}O = ({}^{18}R_{sample}/{}^{18}R_{standard} - 1) \times 1{,}000‰$ with R_{sample} and $R_{standard}$ respectively being the ratios of heavy to light isotopes of the sample and the corresponding standard (atmospheric N_2 for $\delta^{15}N$ and VSMOW for $\delta^{18}O$). To assign $\delta^{15}N$ and $\delta^{18}O$ values to NOAA-1, the recently published N_2O isotopic data spanning the years 1978–2005 retrieved from archived air samples from Cape Grim, Tasmania[31], were linearly extrapolated to the collection date of NOAA-1 (Extended Data Fig. 6). This assignment led to $\delta^{15}N$ and $\delta^{18}O$ values for NOAA-1 of 6.18‰ and 44.16‰, respectively. To test this empirical calibration, a second standard gas cylinder, labelled NOAA-2, filled on 5 October 1988 at Niwot Ridge was used. Its analysis relative to NOAA-1 resulted in respective $\delta^{15}N$ and $\delta^{18}O$ values of $6.93 \pm 0.04‰$ and $44.29 \pm 0.07‰$, and a N_2O concentration of 306.9 ± 0.3 p.p.b. (mean $\pm$ s.e., $n = 8$). The linear interpolation of the Cape Grim data would lead to respective $\delta^{15}N$ and $\delta^{18}O$ values of 6.91‰ and 44.62‰, and a N_2O concentration of 307.1 p.p.b., suggesting a reasonably good agreement within the interannual scatter of the Cape Grim data (Extended Data Fig. 6). As a further test of the calibration scale, and to allow for comparison with future data sets, firn air collected on 27 July 2008 at NEEM, Greenland (dated to approximately AD 1958), was analysed both at Oregon State University and the University of Bern. On average, the $\delta^{15}N$ and $\delta^{18}O$ values measured at Oregon State University were 0.80‰ lower and 0.36‰ higher, respectively, than the values measured at the University of Bern, where an independent primary standard, calibrated with a similar method[32], was used. These differences were again within the range of the interannual scatter observed in the Cape Grim data. Because the isotopic composition of NOAA-1 used as the standard differed from the values measured in the ice samples, a small systematic bias in the absolute values cannot entirely be ruled out, even though the Taylor Glacier data are in good agreement with data from other labs (Extended Data Fig. 2). However, a systematic bias in the reference scale would not affect the conclusions drawn in this manuscript. Finally, an artificial air mixture, labelled NOAA-3, with a N_2O concentration of 283.25 ± 0.09 p.p.b. was available at Oregon State University for additional calibrations and for quality assurance (see below).

Measurement procedure for N_2O concentration and isotopic composition. The analysis of the nitrogen and oxygen isotopic composition of N_2O ($\delta^{15}N$ and $\delta^{18}O$) was performed in a fashion similar to previously described techniques[32–35] using a MAT 253 isotope mass spectrometer in continuous-flow mode, which was coupled to a pre-concentration device and a gas chromatograph. The ancient air was extracted from Taylor Glacier ice samples containing visible air bubbles using two 'cheese grater' devices, that is, electropolished, stainless-steel extraction pots (4.7 l) equipped with perforated, electropolished, stainless-steel plates with sharp edges[36]. Before the ice samples were loaded, the extraction pots were washed with Milli-Q water and ethanol, completely dried at 60 °C (45–60 min) and cooled in a walk-in freezer (about 60 min) to the temperature of the stored ice samples (-25 °C). Taylor Glacier ice samples were cut and cleaned (by removing typically about 200 g of the outermost ice) with a band-saw, resulting in octagonal prisms of about 700–900 g. The loaded extraction pots were sealed with copper gaskets (CF flange), put inside the lab freezer at -60 °C and evacuated for 30 min. The next day, the first extraction pot was evacuated for another 45 min and then the first ice sample was grated by moving the extraction pot back and forth horizontally for one hour in the lab freezer at -60 °C. On average 36% of the ice sample was grated and about 20–40 ml of air was typically extracted. The rather low grating efficiency did not affect the results, because intact ice remained and bubbles were either completely opened or remained closed (the results were also confirmed by intercalibration measurements using a different extraction technique; see below). The air was then expanded into the vacuum system, where traps were installed in the following sequential order: (i) a stainless-steel, 1/4-inch tube forming a spiral at -105 °C, to trap water vapour; (ii) a stainless-steel, 1/4-inch tube forming a double U-trap at liquid-nitrogen temperature, to trap N_2O and CO_2; and (iii) a 1/4-inch cold finger at 11 K, acting as a vacuum pump and trapping the remaining air constituents in about 21 min. By transferring N_2O and CO_2 with a helium flow (47 ml min^{-1}; ultrapure helium additionally cleaned with a hydrocarbon trap, a high-capacity gas purifier and an indicating hydrocarbon, moisture and oxygen trap) through an Ascarite and magnesium perchlorate trap, CO_2 was chemically removed and N_2O was further pre-concentrated in a stainless-steel, 1/16-inch tube forming a U-trap at liquid-nitrogen temperature. N_2O was then transferred with a helium flow (0.9 ml min^{-1})

onto a deactivated, fused-silica capillary (internal diameter, 0.25 mm) immersed in liquid nitrogen that served as a cryofocus. N_2O was separated from remaining traces of CO_2 in a fused-silica gas chromatographic column (Agilent PoraBond Q; internal diameter, 0.32 mm; 25 m) at 24 °C again using a helium flow (0.9 ml min^{-1}). After passing a Nafion dryer, N_2O entered the open split of a Thermo Scientific ConFlo IV and from there the MAT 253 isotope mass spectrometer, where the m/z 44, 45 and 46 beams were monitored. Direct injection of ultrapure N_2O into the open split produced four rectangular peaks of 20 s duration preceding the Gaussian peak eluting from the gas chromatographic column, the latter typically reaching peak areas of 0.7–1.6 V s for m/z 44. Before and after the ice-sample measurement, duplicates of similar chromatograms were produced by repeated pre-concentration of N_2O from an aliquot of NOAA-1 standard gas, resulting in peak areas of 2.5 V s for m/z 44. Furthermore, an additional NOAA-1 standard gas measurement not used for calibration was performed at the beginning to test and condition the measurement system. After baseline correction and peak integration of both the rectangular and Gaussian peaks in each chromatogram using a custom-designed algorithm able to fit exponential baselines, the elemental ratios of each peak were calculated, thereby correcting for the contribution of ^{17}O according to ref. 37. Then the raw $\delta^{15}N$ and $\delta^{18}O$ values of the Gaussian peaks of each chromatogram were determined relative to the mean of the four preceding rectangular peaks. This ensured the removal of any potential drift of the mass spectrometer over the course of the measurement day. Finally, the raw $\delta^{15}N$ and $\delta^{18}O$ values of the ice-sample peak were referenced against the mean raw $\delta^{15}N$ and $\delta^{18}O$ values of the four NOAA-1 standard gas peaks. To determine the N_2O concentration, the air from the ice-sample and NOAA-1 standard gas measurements collected on separate cold fingers was expanded into a previously evacuated stainless-steel cylinder (2.4 l) installed inside the oven of the temperature-stable gas chromatograph, and the pressure was recorded. The N_2O concentration of the ice sample was then calculated by referencing the m/z 44 peak area-to-pressure ratio of the ice-sample measurement against the mean of the m/z 44 peak area-to-pressure ratios of the NOAA-1 standard gas measurements. Finally, the evacuation, grating and measuring procedures were repeated with the second extraction pot to measure two ice samples per day.

Long-term stability, amount dependency and blank ice measurements. To ensure that the data resulting from a measurement series extending over several months were not affected by any systematic drift (for example that caused by unintended changes in the standard gases, pre-concentration system or measurement procedure), NOAA-3 standard gas was analysed daily with peak areas of 1.92 ± 0.05 V s for m/z 44. The results over the full measurement series showed no drift and the standard deviations (1σ, $n = 31$) were 0.14‰ and 0.32‰ for $\delta^{15}N$ and $\delta^{18}O$, respectively (Extended Data Fig. 7a). To investigate the amount dependency of the measurement system over the full range of analysed N_2O amounts (8.5–21.0 ng of N_2O), an aliquot of NOAA-1 standard gas containing the same amount of N_2O as the preceding ice sample was routinely analysed. The N_2O isotopic compositions of these measurements did not show a significant amount dependency (Extended Data Fig. 7b). Their standard deviations (1σ, $n = 58$) were 0.22‰ and 0.59‰ for $\delta^{15}N$ and $\delta^{18}O$, respectively, and were thus somewhat higher than the standard deviations of the NOAA-3 standard gas measurements used to check the stability of the system as mentioned above, probably owing to the smaller peak areas. Ten measurements of different amounts of NOAA-1 standard gas which was stored in the extraction pots while a piece of artificial bubble-free ice was grated further confirmed the absence of a significant amount dependency of $\delta^{15}N$ and $\delta^{18}O$ over the relevant range (Extended Data Fig. 7c). The mean and standard deviation for $\delta^{15}N$ were respectively 6.11‰ and 0.13‰, and, thus, similar to the value of 6.18‰ assigned to NOAA-1. However, the mean and standard deviation of $\delta^{18}O$ were respectively 42.78‰ and 0.70‰, and, thus, 1.38‰ lower than the value assigned to NOAA-1. Although the reasons for this offset remained obscure, the $\delta^{18}O$ data presented in this study were corrected by +1.38‰. The mean N_2O concentration was 323.2 ± 2.8 p.p.b., compared with 322.32 p.p.b. for the NOAA-1 standard gas. According to these results, N_2O concentration data were not corrected for any shift or amount dependency.

Sampling. In the austral summer 2011–2012, ice samples covering the last deglaciation were collected on the Taylor Glacier, Dry Valleys, Antarctica (77° 46′ S, 161° 43′ E), on a horizontal transect perpendicular to the flow line of the glacier ('horizontal core'; D.B. *et al.*, manuscript in preparation). To avoid cracks in the ice caused by thermal stress on the glacier surface, the ice samples were retrieved from a depth of about 4 m. The horizontal distance between ice samples was usually 1 m, and the transect covered a total distance of 276 m including a fold resulting from disturbed ice flow. The timescale was established by synchronization of fast global changes in the Taylor Glacier CH_4 data with the corresponding changes in the WAIS Divide deep ice-core CH_4 data on an updated version of the WDC06A-7 timescale[15]. Between these tie points the synchronization was further constrained by finding the optimal alignment of the Taylor Glacier and WAIS Divide deep ice-core isotopic data of molecular oxygen using a previously described matching technique[38]. On the resulting timescale, which will be presented in detail in forthcoming publications, the timing of major changes in CH_4, CO_2, N_2O and the isotopic composition of

molecular oxygen were in excellent agreement with previous ice-core data from Greenland and Antarctica. For instance, the Taylor Glacier CH_4 data matched the Talos Dome CH_4 data on the independent AICC2012 timescale[26] (Fig. 1). This study focuses on the time interval from 16 to 10 kyr BP because the quality and quantity of collected Taylor Glacier ice samples with greater ages did not allow for reliable N_2O measurements.

Analysis of Taylor Glacier ice samples. Sixty-four Taylor Glacier ice samples were analysed in random order for their N_2O concentrations and $\delta^{15}N$ and $\delta^{18}O$ isotopic compositions. These measurements included ten pairs of replicates, which had pooled standard deviations (1σ) of 3.4 p.p.b. for the N_2O concentration, 0.28‰ for $\delta^{15}N$ and 1.04‰ for $\delta^{18}O$. These standard deviations were considered to represent the best estimates of the measurement uncertainties, because the replicates were affected by most potential sources of disturbance (for example small-scale anomalies in the ice samples, drifts throughout the measurements series, uncertainties associated with baseline correction and peak integration, and so on). The final results were corrected for gravitational enrichment in the firn column on the basis of measurements of the isotopic composition of atmospheric molecular nitrogen using ice samples collected at the same site (D.B. *et al.*, manuscript in preparation). The corrections for all Taylor Glacier ice samples were relatively small and accounted for a reduction of 0.4–1.0 p.p.b. for N_2O, 0.12–0.24‰ for $\delta^{15}N$ and 0.24–0.50‰ for $\delta^{18}O$. Diffusive isotopic fractionation of N_2O in the firn at this site is negligible for the observed concentration increase rates[39].

Effect of atmospheric imbalances on the isotopic composition of tropospheric N_2O. Rapid changes in the atmospheric N_2O load lead to temporal shifts (lasting several lifetimes) in the isotopic composition of tropospheric N_2O even when the overall isotopic composition of the total N_2O source remains unchanged[40]. This is a consequence of the preferred removal of N_2O enriched in light isotopes by the stratospheric sink, which can also be described by a slightly longer atmospheric lifetime of N_2O enriched in heavy isotopes. Accordingly, a hypothetical rapid increase or decrease of global emissions without changing the overall isotopic composition of the emitted N_2O would still temporarily shift tropospheric $\delta^{15}N$ and $\delta^{18}O$ to lighter or, respectively, heavier values. Indeed, the Taylor Glacier data reveal that contemporaneously with the fast increases of the N_2O load from HS1 to the Bølling–Allerød as well as from the Younger Dryas to the Preboral, $\delta^{15}N$ decreased rapidly, whereas it rapidly increased contemporaneously with the decrease of N_2O from the Bølling–Allerød to the Younger Dryas (Fig. 1b). The box model allowed for quantification of the atmospheric imbalances and shows that they explain only part of these observed trends, suggesting that the relative contributions of marine and terrestrial sources changed as well (Fig. 2). This is highlighted by calculation of the hypothetical N_2O concentration and $\delta^{15}N$ isotopic composition which would result when, at any time, the marine and terrestrial emissions reached equilibrium with the sink. The evolution of this equilibrated atmosphere is illustrated with dashed lines in Fig. 2; the differences between dashed and solid lines indicate the effect of atmospheric imbalances accounted for by the box model.

Atmospheric origin of Taylor Glacier N_2O data. Past studies of N_2O on air extracted from polar ice cores were often complicated by *in situ* production of N_2O in the ice matrix, which, for some ice cores, partly contaminated the atmospheric signal[2,3,5,8,33]. Although it is hard to rule out unambiguously any influence of *in situ* production on the new Taylor Glacier concentration and $\delta^{15}N$ and $\delta^{18}O$ data between 16 and 10 kyr BP, the following considerations point to an exclusively atmospheric origin of the reconstructed trends. First, although *in situ* production has been described as occurring randomly with large scatter between nearby ice samples[2,3,5,8,33], the Taylor Glacier N_2O concentration data are smooth and in complete agreement with independent measurements on a second set of ice samples from the same site produced with the apparatus for CO_2 isotopes at Oregon State University[36] (Extended Data Fig. 2), as well as with the atmospheric trends previously reconstructed along various ice cores. However, the fact that the Taylor Glacier ice samples were more than tenfold larger than the samples used for most concentration measurements may have obscured the detection of excess N_2O potentially present owing to *in situ* production on small spatial scales within Taylor Glacier ice samples. Second, the Taylor Glacier $\delta^{15}N$ and $\delta^{18}O$ data between 16 and 10 kyr BP do not show obvious outliers, and replicated measurements showed a satisfactory reproducibility; the standard deviations increased only from 0.22‰ to 0.28‰ for $\delta^{15}N$ and from 0.59‰ to 1.04‰ for $\delta^{18}O$ when measuring natural ice samples instead of standard gas, the latter being unaffected by the extraction process and potential variations in the ice. Third, $\delta^{15}N$ and $\delta^{18}O$ show short-term isotopic excursions partly resulting from imbalances in the emission and removal of N_2O at times of changing atmospheric N_2O load. Because these variations are an expected and well-understood consequence of atmospheric processes related to the preferred removal of N_2O enriched in light isotopes by the stratospheric sink, their presence in $\delta^{15}N$ provides confidence that the Taylor Glacier data indeed represent atmospheric trends.
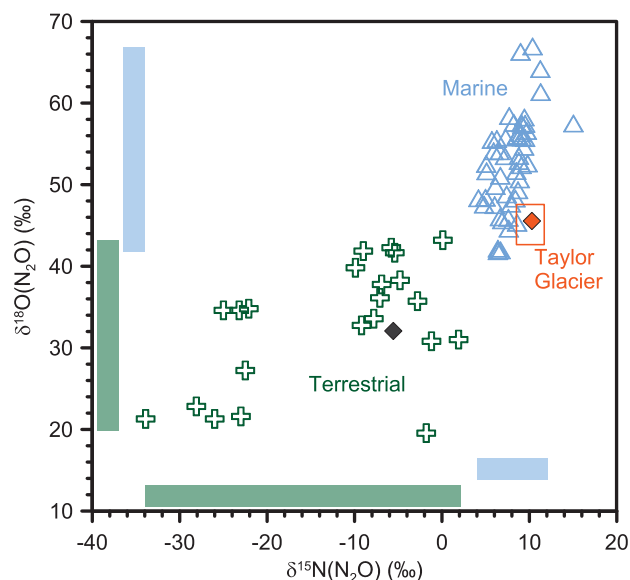
Calculation of marine and terrestrial emissions (two-box model). To calculate the relative contributions of marine and terrestrial sources to total N_2O emissions, a two-box model (similar to refs 9, 10, 41), including a tropospheric and a stratospheric box, a marine and a terrestrial source, as well as a stratospheric sink, was used, with the basic equations shown in Extended Data Table 1. In a Monte Carlo approach, the model parameters (atmospheric lifetime, exchange rate of air between troposphere and stratosphere, stratospheric fractionation constant, and characteristic isotopic compositions of the marine and terrestrial sources) were randomly varied within prescribed distributions, which were based either on measurement uncertainties or the full ranges reported in the literature (Extended Data Table 2). For each random combination of parameters, the marine and terrestrial emissions that reproduce the Taylor Glacier N_2O concentration and $\delta^{15}N$ isotopic data (represented by splines with an empirical cut-off period of 600 yr, which smoothly follow the significant variability in the data) were inversely calculated. Hence, in a forward calculation the determined marine and terrestrial emissions would exactly recover the (splined) Taylor Glacier data. To take into account the measurement uncertainties, for each iteration the Taylor Glacier data were randomly varied within their uncertainties (Gaussian distributions using the pooled standard deviations of ten pairs of replicates), and the splines were re-determined. We argue that on the basis of the available data it is currently not possible to estimate robust global mean values for the isotopic compositions of the marine and terrestrial N_2O sources. Therefore, a conservative approach was used with distributions uniformly covering the full ranges of field data (Extended Data Fig. 1). The Monte Carlo simulations were continued until 500 combinations were found with initial marine fractions (at 16 kyr BP) of 17%, 37% and 74% of total emissions in respective accordance with low, best and high estimates for the modern natural N_2O budget[1] (Extended Data Fig. 5), ignoring all other results (with different initial marine fractions). Using the resulting evolutions of marine and terrestrial emissions to calculate tropospheric $\delta^{18}O$ provided results which were consistent with Taylor Glacier $\delta^{18}O$ for the scenarios with initial marine fractions of 17% and 37% (Extended Data Fig. 4). However, the scenario with an initial marine fraction of 74% was not supported by the isotopic data because it would require $\delta^{18}O$ values of the marine and terrestrial sources outside the range of reported field data (Extended Data Table 2). Finally, to estimate the effect of the marine N_2O cycle and inventory on tropospheric N_2O and $\delta^{15}N$, the two-box model was extended by six stacked ocean boxes. The timescales of exchange between the ocean boxes as well as between the uppermost ocean box and the troposphere were tuned to get the same model response to an instant emission of 200 Tg N into the troposphere as with the Bern3D Earth System Model[42]. Both formulations of the box model produced very similar results, indicating that, owing to the fast exchange of N_2O between the ocean and the atmosphere, any physical effects caused by ocean circulation and N_2O solubility can be neglected for the last deglaciation (Extended Data Fig. 3).

LPX-Bern model. For comparison with the terrestrial N_2O emissions inferred from the box model based on the Taylor Glacier data, terrestrial N_2O emissions over the last deglaciation were independently derived from transient simulations with LPX-Bern, a dynamic global vegetation and land surface process model. We applied the most recently published version of the model[11] (v1.0), with input data and set-up as published in ref. 43. The LPX-Bern model describes dynamical vegetation and terrestrial biogeochemical processes, and integrates representations of non-peatland[44–46] and peatland[43,47,48] ecosystems and their carbon and nitrogen dynamics[11,49,50]. The model calculates the release and uptake of the trace gases CO_2, N_2O (refs 11, 49, 50) and CH_4 (refs 51–53). Plant functional types (PFTs) are the basic biological units and represent different life forms (grasses, trees, mosses) and combinations of plant traits (needle-leaved, broad-leaved and so on). These PFTs are in competition for resources (water, light, nitrogen) on each grid cell and land unit (for example peat and non-peat). The model accounts for the coupling of carbon and water cycles through photosynthesis and evapotranspiration. It uses a vertically resolved soil hydrology, heat diffusion and an interactive thawing–freezing scheme[46,47]. The LPX-Bern vegetation component interacts with a dynamic nitrogen-cycle module that includes the relevant nitrogen fluxes and pools for plants and soils. The nitrogen source is implied by keeping the ratio of soil carbon to nitrogen constant over time. Thus, in LPX-Bern plant growth is not directly limited by external nitrogen input into an ecosystem, but by the rate of nitrogen remineralisation for a given climate. The total global N_2O emissions in LPX-Bern depend on the model parameters for the nitrogen fraction emitted as N_2O during denitrification and the fraction of nitrogen leaching in the form of N_2O in runoff. Although changes in soil texture over time could have an impact on terrestrial N_2O emissions[54], no information on soil texture changes over the last deglaciation is available and the modern field was applied, representing a potential source of uncertainty. The input climate (temperature, precipitation, cloud cover, wet days) was obtained from anomalies of transient climate simulations over the past 21 kyr with the NCAR CCSM3 (TraCE-21ka[25,55]) and observed present day climate (CRU[56]). Further input data were atmospheric CO_2 (ref. 57), orbital insolation changes[58] and topography changes through ice-sheet and sea-level changes imposed by ICE-5G[59]. Here the LPX-Bern model was run with a spatial resolution of $3.75° \times 2.5°$ and a daily time step was applied in the photosynthesis, water and nitrogen modules. Simulations started from an equilibrated spin-up at 21 kyr BP. Note that as we were

applying LPX-Bern as used in the literature, the absolute emissions at 16 kyr BP were 6.2 Tg N yr^{-1}, whereas the value corresponding to an initial marine fraction of 37% of the total source is 4.6 Tg N yr^{-1}. Several potential biases might explain differences in the magnitude of terrestrial N_2O emissions and emission changes inferred from LPX-Bern and the Taylor Glacier data: the initial terrestrial fraction (at 16 kyr BP) in the box model could be overestimated (Methods and Extended Data Fig. 5), the sensitivity of LPX-Bern to temperature changes could be too low (LPX-Bern shows a positive sensitivity to changes in temperature), or the temperature anomalies from TraCE-21ka could be damped relative to climate during the last deglaciation (TraCE-21ka indeed shows a relatively modest warming over the last deglaciation and modest changes associated with the Younger Dryas). However, the LPX-Bern simulations show reasonable quantitative emissions without any further tuning, and are used here as an independent approach to further support the variability in terrestrial N_2O emissions as inferred from the Taylor Glacier data.

31. Park, S. *et al.* Trends and seasonal cycles in the isotopic composition of nitrous oxide since 1940. *Nature Geosci.* **5,** 261–265 (2012).
32. Schmitt, J., Seth, B., Bock, M. & Fischer, H. Online technique for isotope and mixing ratios of CH_4, N_2O, Xe and mixing ratios of organic trace gases on a single ice core sample. *Atmos. Meas. Tech.* **7,** 2645–2665 (2014).
33. Sowers, T. N_2O record spanning the penultimate deglaciation from the Vostok ice core. *J. Geophys. Res.* **106,** 31903–31914 (2001).
34. Sapart, C. J. *et al.* Simultaneous stable isotope analysis of methane and nitrous oxide on ice core samples. *Atmos. Meas. Tech.* **4,** 2607–2618 (2011).
35. Sperlich, P. *et al.* An automated GC-C-GC-IRMS setup to measure palaeoatmospheric $\delta^{13}C$-CH_4, $\delta^{15}N$-N_2O and $\delta^{18}O$-N_2O in one ice core sample. *Atmos. Meas. Tech.* **6,** 2027–2041 (2013).
36. Bauska, T. K., Brook, E. J., Mix, A. C. & Ross, A. High-precision dual-inlet IRMS measurements of the stable isotopes of CO_2 and the N_2O/CO_2 ratio from polar ice core samples. *Atmos. Meas. Tech.* **7,** 3825–3837 (2014).
37. Kaiser, J. *Stable Isotope Investigations of Atmospheric Nitrous Oxide* 17–21. PhD thesis, Johannes Gutenberg Univ. Mainz (2002).
38. Lisiecki, L. E. & Lisiecki, P. A. Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography* **17,** 1049 (2002).
39. Buizert, C., Sowers, T. & Blunier, T. Assessment of diffusive isotopic fractionation in polar firn, and application to ice core trace gas records. *Earth Planet. Sci. Lett.* **361,** 110–119 (2013).
40. Tans, P. P. A note on isotopic ratios and the global atmospheric methane budget. *Glob. Biogeochem. Cycles* **11,** 77–81 (1997).
41. Sowers, T., Rodebaugh, A., Yoshida, N. & Toyoda, S. Extending records of the isotopic composition of atmospheric N_2O back to 1800 AD from air trapped in snow at the South Pole and the Greenland Ice Sheet Project II ice core. *Glob. Biogeochem. Cycles* **16,** 1129 (2002).
42. Roth, R. *Modeling Forcings and Responses in the Global Carbon Cycle-Climate System: Past, Present and Future.* PhD thesis, Univ. Bern (2013).
43. Spahni, R. *et al.* Transient simulations of the carbon and nitrogen dynamics in northern peatlands: from the Last Glacial Maximum to the 21st century. *Clim. Past* **9,** 1287–1308 (2013).
44. Sitch, S. *et al.* Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* **9,** 161–185 (2003).
45. Joos, F. *et al.* Transient simulations of Holocene atmospheric carbon dioxide and terrestrial carbon since the Last Glacial Maximum. *Glob. Biogeochem. Cycles* **18,** GB2002 (2004).
46. Gerten, D. *et al.* Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model. *J. Hydrol. (Amst.)* **286,** 249–270 (2004).
47. Wania, R., Ross, I. & Prentice, I. C. Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes. *Glob. Biogeochem. Cycles* **23,** GB3014 (2009).
48. Wania, R., Ross, I. & Prentice, I. C. Integrating peatlands and permafrost into a dynamic global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle processes. *Glob. Biogeochem. Cycles* **23,** GB3015 (2009).
49. Xu-Ri & Prentice, I. C. Terrestrial nitrogen cycle simulation with a dynamic global vegetation model. *Glob. Change Biol.* **14,** 1745–1764 (2008).
50. Xu-Ri, Prentice I. C., Spahni, R. & Niu, H. S. Modelling terrestrial nitrous oxide emissions and implications for climate feedback. *New Phytol.* **196,** 472–488 (2012).
51. Wania, R., Ross, I. & Prentice, I. C. Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1. *Geosci. Model Dev.* **3,** 565–584 (2010).
52. Spahni, R. *et al.* Constraining global methane emissions and uptake by ecosystems. *Biogeosciences* **8,** 1643–1665 (2011).
53. Zürcher, S. *et al.* Impact of an abrupt cooling event on interglacial methane emissions in northern peatlands. *Biogeosciences* **10,** 1963–1981 (2013).
54. Pfeiffer, M., van Leeuwen, J., van der Knaap, W. O. & Kaplan, J. O. The effect of abrupt climatic warming on biogeochemical cycling and N_2O emissions in a terrestrial ecosystem. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **391,** 74–83 (2013).
55. He, F. *Simulating Transient Climate Evolution of the Last Deglaciation with CCSM3.* PhD thesis, Univ. Wisconsin-Madison (2011).
56. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25,** 693–712 (2005).
57. Joos, F. & Spahni, R. Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years. *Proc. Natl Acad. Sci. USA* **105,** 1425–1430 (2008).
58. Berger, A. L. Long-term variations of daily insolation and quaternary climatic changes. *J. Atmos. Sci.* **35,** 2362–2367 (1978).
59. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: The ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32,** 111–149 (2004).
60. Kim, K.-R. & Craig, H. Two-isotope characterization of N_2O in the Pacific Ocean and constraints on its origin in deep water. *Nature* **347,** 58–61 (1990).
61. Yoshinari, T. *et al.* Nitrogen and oxygen isotopic composition of N_2O from suboxic waters of the eastern tropical North Pacific and the Arabian Sea—measurement by continuous-flow isotope-ratio monitoring. *Mar. Chem.* **56,** 253–264 (1997).
62. Kim, K.-R. & Craig, H. Nitrogen-15 and oxygen-18 characteristics of nitrous oxide: a global perspective. *Science* **262,** 1855–1857 (1993).
63. Pérez, T. *et al.* Isotopic variability of N_2O emissions from tropical forest soils. *Glob. Biogeochem. Cycles* **14,** 525–535 (2000).
64. Rahn, T. & Wahlen, M. Stable isotope enrichment in stratospheric nitrous oxide. *Science* **278,** 1776–1778 (1997).
65. Röckmann, T. *et al.* Isotopic enrichment of nitrous oxide ($^{15}N^{14}NO$, $^{14}N^{15}NO$, $^{14}N^{14}N^{18}O$) in the stratosphere and in the laboratory. *J. Geophys. Res.* **106,** 10403–10410 (2001).

Extended Data Figure 1 | Isotopic composition of marine and terrestrial N$_2$O sources. Field data of δ^{15}N (relative to atmospheric N$_2$) and δ^{18}O (relative to VSMOW) of marine (blue triangles[60,61]) and terrestrial (green crosses[62,63]) N$_2$O sources. Blue and green bars indicate the ranges as used in the box model (Extended Data Table 2). The mean tropospheric value of all Taylor Glacier data (orange diamond, with the orange box indicating the full range of the data) is enriched in heavy isotopes in both δ^{15}N and δ^{18}O relative to the approximate corresponding isotopic composition of the total source (black diamond) owing to the fractionation by the stratospheric sink.
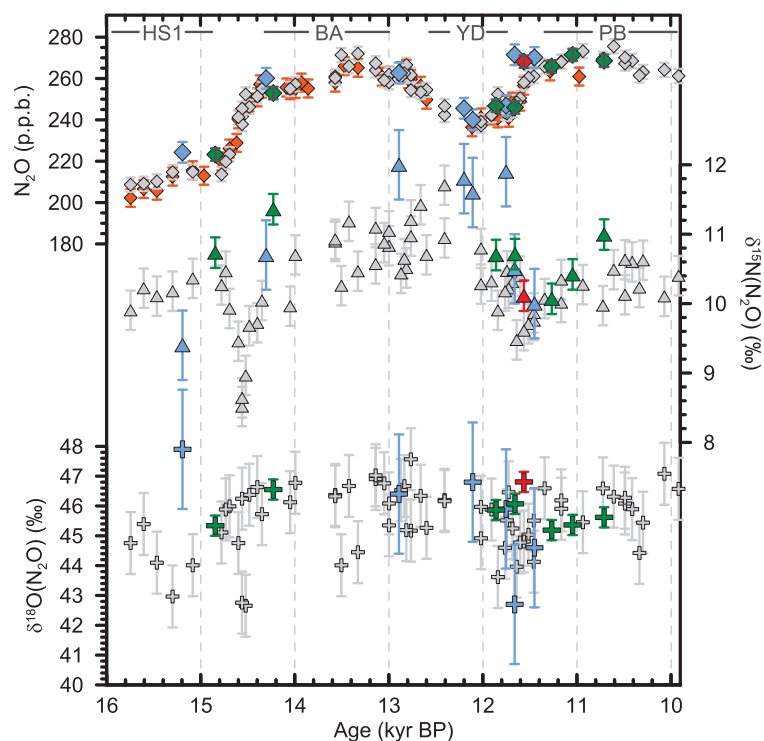
Extended Data Figure 2 | Comparison of Taylor Glacier and other data. N_2O concentration (diamonds), $\delta^{15}N$ (triangles) and $\delta^{18}O$ (crosses) data from Taylor Glacier (from the apparatus for N_2O isotopes in grey, and from the apparatus for CO_2 isotopes[36] in orange) compared with a Taylor Glacier intercomparison measurement (red) and Talos Dome data (green) from the University of Bern, and with published data from Taylor Dome (blue[4]). Taylor Glacier and Talos Dome data from the University of Bern were corrected by +2.28 p.p.b. for the N_2O concentration, $-0.80‰$ for $\delta^{15}N$ and $+0.36‰$ for $\delta^{18}O$ on the basis of intercalibration measurements made by Oregon State University and the University of Bern using firn air (Methods). The Taylor Glacier $\delta^{18}O$ data from Oregon State University (grey crosses) were corrected by $+1.38‰$ on the basis of measurements with bubble-free ice and NOAA-1 standard gas (Extended Data Fig. 7c and Methods). Error bars indicate $\pm 1\sigma$.
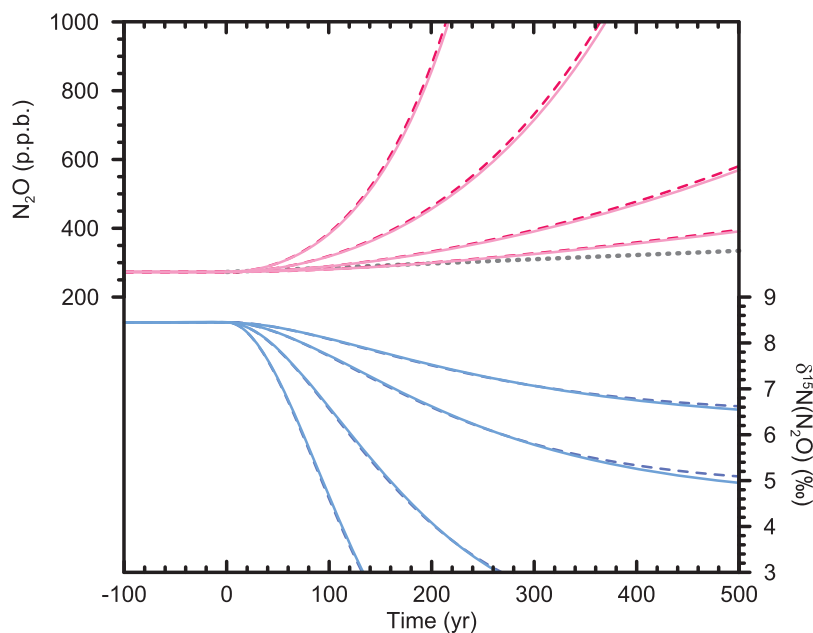
Extended Data Figure 3 | Effect of marine N$_2$O cycle and inventory on tropospheric N$_2$O concentration and δ^{15}N under changing emissions. Response of tropospheric N$_2$O and δ^{15}N to exponential increases in N$_2$O emissions with timescales of 100, 200, 500 and 1,000 yr (from left to right; the maximum increase rate in the Taylor Glacier N$_2$O data is indicated by the dotted grey line). Results from the two-box model (dashed lines show results without explicit representation of the ocean) are very similar to the results from the extended box model (with ocean, solid lines; Methods). The marine and terrestrial emissions are increased in parallel, that is, the marine fraction is always 37% of the total N$_2$O emissions and the isotopic composition of emitted N$_2$O remains constant. The decrease in δ^{15}N is caused by imbalances between the sources and the stratospheric sink.
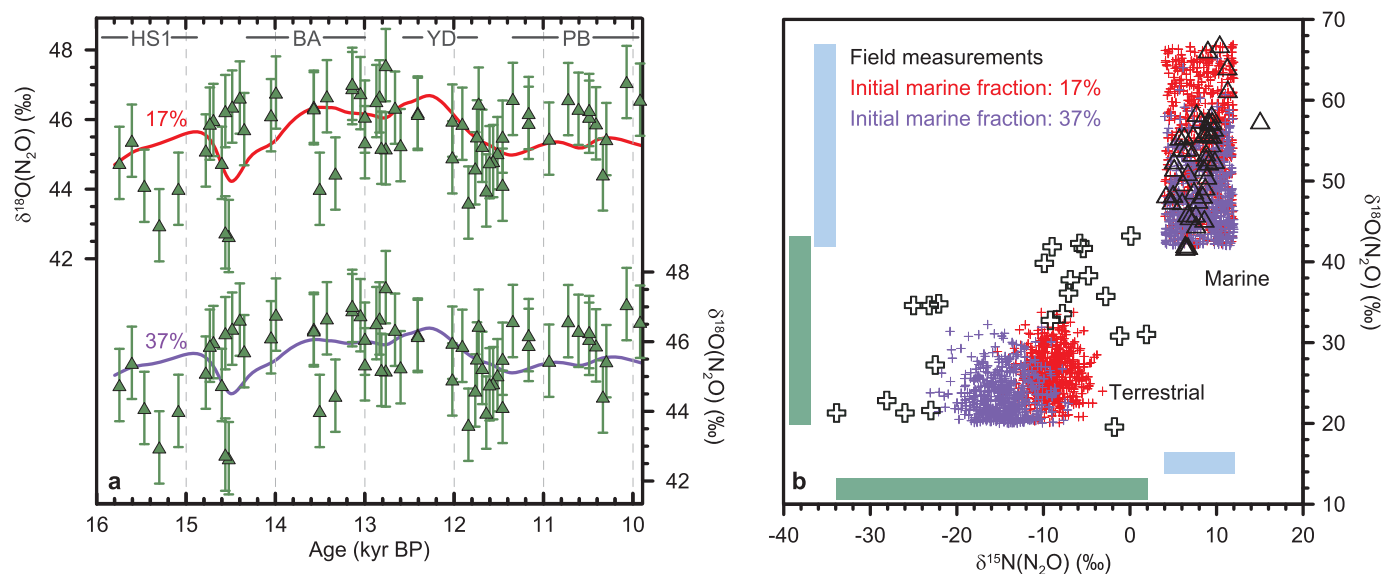
Extended Data Figure 4 | Consistency of the calculated marine and terrestrial emissions with the Taylor Glacier $\delta^{18}O$ data. a, $\delta^{18}O$ evolution for different initial marine fractions (red, 17%; purple, 37%) of the total emissions when calculated using the marine and terrestrial N_2O emissions determined on the basis of the Taylor Glacier N_2O concentration and $\delta^{15}N$ data. In a Monte Carlo approach only scenarios with the same mean value as the Taylor Glacier $\delta^{18}O$ data (green, with $\pm 1\sigma$ error bars) were considered, which narrowed the possible $\delta^{18}O$ isotopic composition of the sources. **b,** $\delta^{15}N$ and $\delta^{18}O$ of marine (triangles) and terrestrial (crosses) sources from modern field data (black, as in Extended Data Fig. 1), as well as the values needed to explain the Taylor Glacier data with different initial (at 16 kyr BP) marine fractions (red, 17%; purple, 37%). An initial marine fraction of 74% would require $\delta^{18}O$ isotopic compositions outside the observed range (Extended Data Table 2), suggesting that such a high marine fraction is rather unlikely. Note that the Taylor Glacier data can be explained for an initial marine fraction of 74% when considering $\delta^{15}N$ only, but only with rather extreme model parameters (Extended Data Fig. 5).
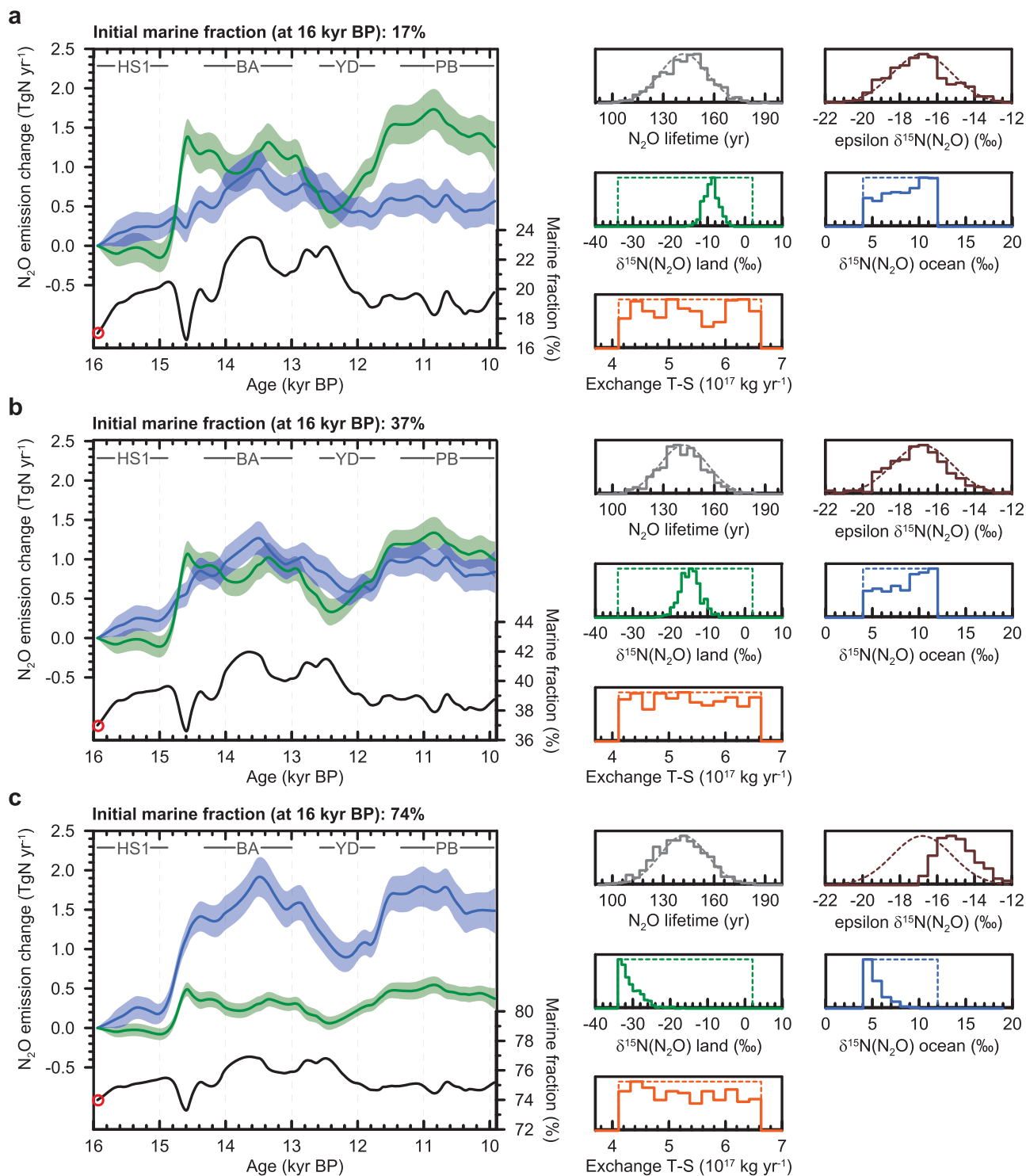
Extended Data Figure 5 | Evolution of marine and terrestrial N$_2$O emissions under different scenarios. Sensitivity of marine (blue in large panels) and terrestrial (green in large panels) N$_2$O emissions to initial marine fractions (red circles at 16 kyr BP) set to 17% (**a**), 37% (**b**) and 74% (**c**) in accordance with low, best and high estimates of the modern natural N$_2$O budget[1]. The uncertainty bands related to the emissions (blue and green shaded areas) result from the Monte Carlo approach and indicate ±1σ of all solutions.

For all scenarios, the maximum absolute changes in the marine fractions (black in large panels) over the last deglaciation are 7% or less. Dashed lines in the small panels show the distributions of the parameters as allowed for in the Monte Carlo approach (priors; Extended Data Table 2), and solid lines indicate the distributions of the parameters which allow for a reproduction of the Taylor Glacier N$_2$O concentration and δ^{15}N that respects the prescribed initial marine fractions (posteriors).
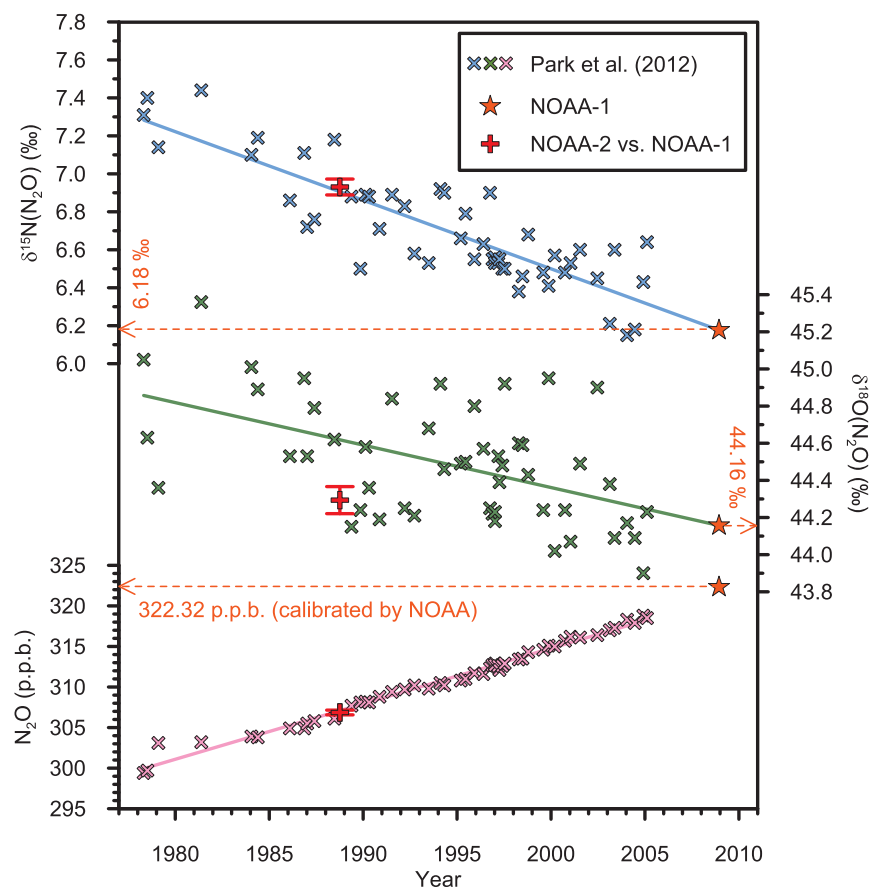
Extended Data Figure 6 | Standard gases for $\delta^{15}N$, $\delta^{18}O$ and N_2O concentration. The $\delta^{15}N$ and $\delta^{18}O$ values of 6.18‰ and 44.16‰ of the NOAA-1 standard gas collected at Niwot Ridge, Colorado, were assigned by linear extrapolation of the data from Cape Grim, Tasmania[31], to the collection date of NOAA-1 (11 December 2008), on the basis of the assumption that N_2O and its isotopes are well mixed in the troposphere owing to the rather long atmospheric lifetime. The N_2O concentration of 322.32 ± 0.14 p.p.b. of the NOAA-1 standard gas was directly determined by the National Oceanic and Atmospheric Administration (NOAA-2006A scale; linear extrapolation of the Cape Grim data would lead to 320.9 p.p.b.). To test the calibrations for the N_2O concentration and isotopic compositions, a second standard gas, NOAA-2, which was collected at Niwot Ridge on 5 October 1988, was measured against NOAA-1. The results (red crosses) were in good agreement with the Cape Grim data, in particular when taking into account the interannual scatter observed in the archived air.
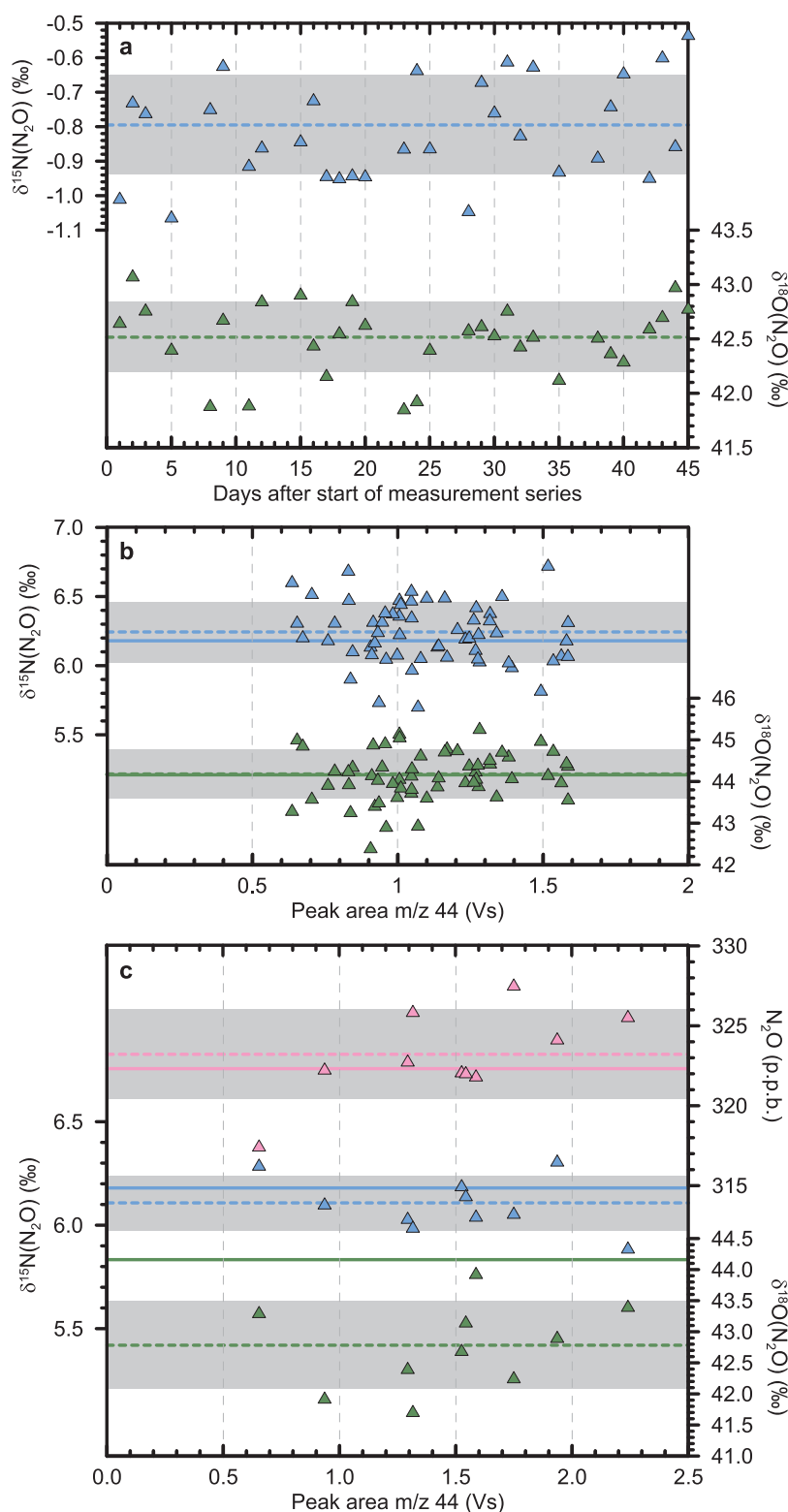
Extended Data Figure 7 | Stability in the course of the measurement series, characterization of the amount dependency of the measurement system, and tests with bubble-free ice. a, NOAA-3 standard gas measurements performed at the end of each measurement day. No significant drifts were observed in the course of the measurement series, and the standard deviations for $\delta^{15}N$ and $\delta^{18}O$ were respectively 0.14‰ and 0.32‰ ($n = 31$), as indicated by the grey areas ($\pm 1\sigma$) around the means (dashed lines). **b**, NOAA-1 standard gas measurements resulting in similar peak areas to the preceding ice-sample measurement routinely performed throughout the measurement series. These measurements covering the full range of peak areas from ice samples did not reveal any significant amount dependency. The mean and standard deviation

(dashed lines and grey areas) for $\delta^{15}N$ were 6.24‰ $\pm$ 0.22‰ and those for $\delta^{18}O$ were 44.18‰ $\pm$ 0.59‰ ($\pm 1\sigma$, $n = 58$), in agreement with the expected values (solid lines). **c**, Measurements of different amounts of NOAA-1 standard gas which was stored in the extraction pots while pieces of bubble-free ice were grated. Dashed lines and grey areas indicate the means and standard deviations ($\pm 1\sigma$, $n = 10$), and solid lines indicate the expected values. On the basis of these measurements with bubble-free ice and the results shown in **b**, N_2O, $\delta^{15}N$ and $\delta^{18}O$ were not corrected for amount dependency. However, during the extraction of standard gas over ice, a -1.38‰ offset was introduced in $\delta^{18}O$. Accordingly, all Taylor Glacier $\delta^{18}O$ values were corrected by $+1.38$‰.

Extended Data Table 1 | Equations forming the basis of the two-box model used to calculate marine and terrestrial N$_2$O emissions

$$\frac{dM_{N_2O,trop}}{dt} = \underbrace{F_{ocean} + F_{land}}_{F_{total}} + Ex_{trop,strat}\big([N_2O]_{strat} - [N_2O]_{trop}\big) \tag{1}$$

$$\frac{dM_{N_2O,strat}}{dt} = Ex_{trop,strat}\big([N_2O]_{trop} - [N_2O]_{strat}\big) - F_{sink} \tag{2}$$

$$F_{sink} = \frac{M_{N_2O,strat}}{k_{strat}} \quad \text{with} \quad k_{strat} = \tau_{N_2O}\chi_{strat} - \frac{M_{atm}}{Ex_{trop,strat}}\chi_{strat}(1 - \chi_{strat}) \tag{3}$$

$$\frac{d(M_{N_2O,trop}R_{trop})}{dt} = fF_{total}R_{ocean} + (1-f)F_{total}R_{land} + Ex_{trop,strat}\big([N_2O]_{strat}R_{strat} - [N_2O]_{trop}R_{trop}\big) \tag{4}$$

$$\frac{d(M_{N_2O,strat}R_{strat})}{dt} = Ex_{trop,strat}\big([N_2O]_{trop}R_{trop} - [N_2O]_{strat}R_{strat}\big) - F_{sink}R_{strat}\Big(\frac{\varepsilon}{1000} + 1\Big) \tag{5}$$

Parameters (see also Extended Data Table 2): M_{atm}, the total mass of the atmosphere; χ_{strat}, the stratospheric fraction of the total atmosphere; $M_{N_2O,trop}$ and $M_{N_2O,strat}$, the masses of N$_2$O in the troposphere and stratosphere; F_{ocean} and F_{land}, the marine and terrestrial N$_2$O emissions; $Ex_{trop,strat}$, the rate of exchange of air between troposphere and stratosphere; $[N_2O]_{trop}$ and $[N_2O]_{strat}$, the N$_2$O mass concentrations in the troposphere and stratosphere; F_{sink}, the rate of removal of N$_2$O from the stratosphere; τ_{N_2O}, the atmospheric lifetime of N$_2$O in equilibrium; k_{strat}, the stratospheric lifetime of N$_2$O; f and $1 - f$, the marine and terrestrial fractions of total N$_2$O emissions; ε, the fractionation constant of the stratospheric sink as defined in ref. 64; R_{trop} and R_{strat}, the isotopic ratios in the troposphere and in the stratosphere; R_{ocean} and R_{land}, the isotopic ratios of marine and terrestrial emissions.

Extended Data Table 2 | Parameters for the two-box model used to calculate marine and terrestrial N₂O emissions

Parameter	Range	Unit	Distribution	Reference
Lifetime	142±14	yr	Gaussian	Ref. 12
$\delta^{15}N(N_2O)$ marine	[4; 12]	‰	Uniform	See Extended Data Figure 1
$\delta^{15}N(N_2O)$ terrestrial	[-34; 2]	‰	Uniform	See Extended Data Figure 1
$\delta^{18}O(N_2O)$ marine	[42; 67]	‰	Uniform	See Extended Data Figure 1
$\delta^{18}O(N_2O)$ terrestrial	[20; 43]	‰	Uniform	See Extended Data Figure 1
Stratospheric fractionation constant (^{15}N)	-16.8±1.6	‰	Gaussian	Ref. 65
Stratospheric fractionation constant (^{18}O)	-13.8±2.0	‰	Gaussian	Ref. 65
Number of moles in atmosphere	1.77×10^{20}	mol	Constant	As in ref. 10
Stratospheric fraction of total atmosphere	0.15	-	Constant	As in ref. 10
Exchange rate troposphere/stratosphere	$[4.11 \times 10^{17}; 6.63 \times 10^{17}]$	kg yr^{-1}	Uniform	As in ref. 9

LETTER

Cell differentiation and germ–soma separation in Ediacaran animal embryo-like fossils

Lei Chen[1,2], Shuhai Xiao[3], Ke Pang[1,2], Chuanming Zhou[1] & Xunlai Yuan[1,4]

Phosphorites of the Ediacaran Doushantuo Formation (∼600 million years old) yield spheroidal microfossils with a palintomic cell cleavage pattern[1,2]. These fossils have been variously interpreted as sulphur-oxidizing bacteria[3], unicellular protists[4], mesomycetozoean-like holozoans[5], green algae akin to *Volvox*[6,7], and blastula embryos of early metazoans[1,2,8–10] or bilaterian animals[11,12]. However, their complete life cycle is unknown and it is uncertain whether they had a cellularly differentiated ontogenetic stage, making it difficult to test their various phylogenetic interpretations. Here we describe new spheroidal fossils from black phosphorites of the Doushantuo Formation that have been overlooked in previous studies. These fossils represent later developmental stages of previously published blastula-like fossils, and they show evidence for cell differentiation, germ–soma separation, and programmed cell death. Their complex multicellularity is inconsistent with a phylogenetic affinity with bacteria, unicellular protists, or mesomycetozoean-like holozoans. Available evidence also indicates that the Doushantuo fossils are unlikely crown-group animals or volvocine green algae. We conclude that an affinity with cellularly differentiated multicellular eukaryotes, including stem-group animals or algae, is likely but more data are needed to constrain further the exact phylogenetic affinity of the Doushantuo fossils.

The Ediacaran Doushantuo Formation at Weng'an in South China provides a valuable taphonomic window into the early evolution of complex multicellular eukaryotes, including florideophyte red algae[13] and possible animals[1]. Among the most controversial and potentially most important Doushantuo fossils is the spheroidal *Megasphaera*, with one or more cells enclosed in a thick ornamented envelope[2,14]. *Megasphaera* exhibits a pattern of palintomic cell cleavage—rapid cell divisions without cytoplasmic growth in between, thus resulting in exponential cell volume decrease with doubling cell number. The first few divisions produce 2–64 tightly sutured polyhedral cells in *Parapandorina*-stage fossils, and subsequent divisions give rise to hundreds to thousands of somewhat loosely packed spherical cells in *Megaclonophycus*-stage fossils. *Megasphaera* has been variously interpreted as sulphur-oxidizing bacteria[3], unicellular protists[4], mesomycetozoean-like holozoans[5], *Volvox*-like green algae[6,7], and blastula embryos of metazoans[1,2,8–10] or bilaterian animals[11,12]. One of the difficulties facing the phylogenetic interpretation of *Megasphaera* lies in its poorly understood life cycle with phylogenetically non-diagnostic characters. For example, although palintomic cell division is universal in early cleavages of animal embryos, it can also occur in gonidial embryos of some *Volvox* species[15] and apparently also in vacuolated cells of the sulphur bacterium *Thiomargarita*[3] and some mesomycetozoeans[5,16]. The ornamented envelopes of *Megasphaera* and related microfossils can be similar in ultrastructure to some animal diapause eggs[10], but such similarities may be convergent[5]. It has been inferred, on the basis of cell division topologies with tightly sutured polyhedral cells meeting at Y-shaped junctions, that *Parapandorina*-stage cells may have lacked a rigid cell wall and were held together by cell adhesion proteins[17,18]. However, the lack of fossils at more advanced and cellularly differentiated ontogenetic stages represents a major objection to the animal interpretation. Thus, not only would the discovery of cellularly differentiated ontogenetic stages lead to a more complete understanding of the life cycle of *Megasphaera*, but it could also distinguish its various phylogenetic interpretations: the bacterial, unicellular protist, and mesomycetozoean-like holozoan interpretations predict that the *Megasphaera* life cycle did not have a multicellular stage with spatial cellular differentiation, but the *Volvox*-like green algal and animal interpretations predict that it did.

Our investigation of the Doushantuo Formation at Weng'an, focusing on analysis of thin sections of the previously overlooked black phosphorite, recovered several cellularly differentiated spheroidal fossils, as well as microfossils representing early developmental stages of *Megasphaera*. Specimens at one-cell, *Parapandorina*-, and *Megaclonophycus*-stage (Fig. 1) are similar to previously published material. Their cell size decreases exponentially as cell number doubles, consistent with palintomic division (Extended Data Fig. 1a, b). Additionally, well-preserved specimens are enclosed by envelopes bearing tuberculate and conical ornaments (black arrowheads in Figs 1a, c, e and 2d). A few *Megaclonophycus*-stage specimens contain tightly sutured polyhedral cells (Fig. 1c), similar to the closely packed polyhedral cells of *Parapandorina*-stage specimens, but different from previously published *Megaclonophycus*-stage specimens with spherical cells in tangential contact or loose aggregation (see, for example, Fig. 1g, h). Furthermore, some *Megaclonophycus*-stage specimens seem to have a peripheral layer of cells that are different in size, shape, and arrangement from the interior cells (black arrows Fig. 1d–f). Taphonomic and diagenetic artefacts are common in these fossils. For example, clear isopachous cement surrounding a carbonaceous phosphatic mass (white arrowheads in Fig. 1a, b) is interpreted as secondary taphonomic structures whose morphologies have no biological meaning[19–21], and intracellular structures (white arrows in Fig. 1c, e, g, h) may also be late diagenetic structures[22].

A group of *Megaclonophycus*-like fossils in our collection contain blastomere-like monad cells, as well as dyad and tetrad cell packets (Fig. 2). These fossils are similar to *Megaclonophycus*-stage fossils in that their monad cell size follows the trend of palintomic cell division (Extended Data Fig. 1a, b). Their cells and cell packets can be tightly (Fig. 2a) or loosely (Fig. 2b, d–g) packed within the enclosing envelope. The dyads contain two hemispherical cells adpressed against each other and sometimes appear to be surrounded by a common membrane (Fig. 2c). The tetrads (40–60 μm in packet size and 20–30 μm in cell size) can be tetrahedral (Fig. 2a) or cruciate (Fig. 2h). Monads, dyads, and tetrads can co-exist in the same specimen (Fig. 2), indicating that cell division is not strictly synchronous. In a few specimens, dyads are concentrated in the interior, whereas slightly elongate monads are arranged in a palisade along the peripheral margin (Fig. 2a), intriguingly resembling the peripheral cell layer in *Megaclonophycus*-stage fossils (Fig. 1d–f). If proved to be a biological feature, the peripheral cell layer is indicative of some degree of cellular differentiation.

Clear evidence of cellular differentiation comes from another group of *Megaclonophycus*-like fossils (Fig. 3). These fossils also follow the

[1]State Key Laboratory of Palaeobiology and Stratigraphy, and Key Laboratory of Economic Stratigraphy and Palaeogeography, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China. [2]College of Earth Science, University of Chinese Academy of Sciences, Beijing 100049, China. [3]Department of Geosciences, Virginia Tech, Blacksburg, Virginia 24061, USA. [4]Department of Geology, Northwest University, Xi'an 710069, China.
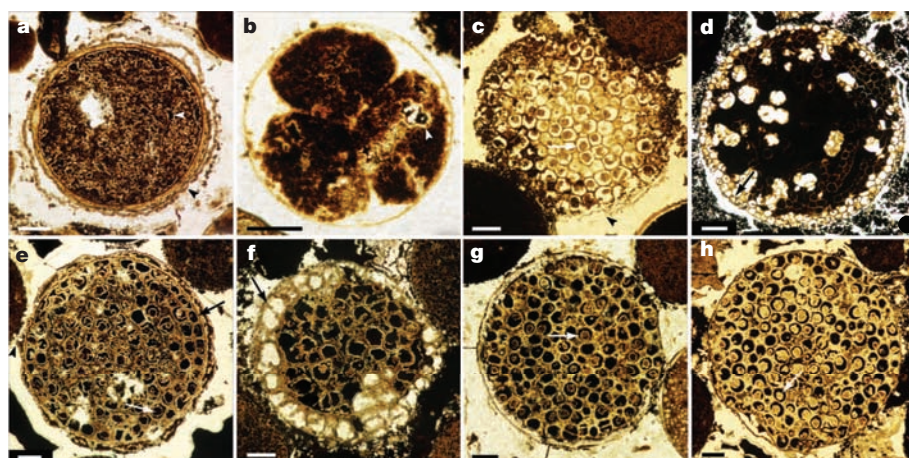
Figure 1 | Early developmental stages of *Megasphaera*. a, One-cell stage. **b**, *Parapandorina*-stage. **c–h**, *Megaclonophycus*-stage. Note tightly sutured polyhedral cells in **b** and **c**, a possible peripheral cell layer in **d–f**, and somewhat loosely aggregated spherical cells in **g–h**. Black arrowheads: tubercular or conical ornamentation on envelopes; white arrowheads: isopachous cement; black arrows: peripheral cell layer; white arrows: nucleus-like diagenetic structures. Scale bars: 100 μm.

trend of palintomic cell division established by *Parapandorina*- and *Megaclonophycus*-stage fossils (Extended Data Fig. 1a, b). However, in addition to the blastomere-like cells, these fossils also contain one or more spheroidal to ellipsoidal multicellular structures, here termed matryoshkas in reference to their similarity to nested Russian dolls. The matryoshkas are of variable size (30–350 μm) but they are generally larger than blastomere-like cells. They themselves are multicellular, consisting of tightly packed cells (9–14 μm in size) that are significantly smaller than the blastomere-like cells. Measurements show that the matryoshkas do not follow a palintomic cell division pattern: cell size does not vary systematically with cell number or matryoshka size; instead, it remains constant, and cell number determines matryoshka size (Extended Data Fig. 1c, d). Thus, matryoshkas are growing structures, with cytoplasmic growth after each division to restore cell size.

We interpret the tightly sutured polyhedral cells (Fig. 1c) as biological features, and the loose aggregation of spherical cells (Fig. 1g, h) as a degradational artefact. This interpretation is based on the unlikelihood of taphonomic transformation from loosely aggregated spherical cells to tightly sutured polyhedral cells. It is further supported by degradation experiments showing that, during initial degradation, germ layer cells of sea urchin embryos are loosened, rounded, and disaggregated because of the degradation of cell adhesion proteins[23]. Thus, *Megaclonophycus*-stage specimens with tightly sutured polyhedral cells (Fig. 1c) suggest that cell adhesion is a biological feature in both *Parapandorina*- and *Megaclonophycus*-stages[17,18].

The new fossils with dyads, tetrads, and matryoshkas are interpreted as later developmental stages of *Megaclonophycus*-stage fossils. They are similar in size and envelope ornamentation, and their constituent monad cells follow the same palintomic cell division pattern as in *Parapandorina*- and *Megaclonophycus*-stage fossils (Extended Data Fig. 1b). This interpretation implies that the dyads, tetrads, and matryoshkas are cell division

products of blastomere-like monads. The developmental continuation from monads to matryoshkas indicates that the matryoshkas are indigenous to the spheroidal fossils, rather than exotic parasites, symbionts, or saprophytes. Because only a small number of monads eventually develop into matryoshkas, this is clear evidence of spatial cellular differentiation. The matryoshkas probably served as an asexual reproduction structure akin to the gonidial embryos of *Volvox carteri*. If so, reproductive cells were sequestered and germ cells were separated from somatic cells by the matryoshka-stage. Insofar as the enclosing envelope maintains a constant size, matryoshka growth must have been accommodated by programmed cell death of somatic cells.

Germ–soma separation represents a form of spatial cell differentiation[24] where altruistic somatic cells undergo programmed cell death to support the reproductive success of germ cells. This altruistic behaviour is probably related to kin selection in early colonial or multicellular organisms[25], and it is a key step towards complex multicellularity[26,27]. This degree of spatial cell differentiation—with inferred presence of cell-to-cell adhesion[18], cell differentiation, germ–soma separation, and programmed cell death—is unknown in modern mesomycetozoeans, unicellular protists, or sulphur-oxidizing bacteria. Thus, phylogenetic interpretations of the Doushantuo fossils based on these modern analogues are questionable.

The striking similarity between the matryoshkas and asexual gonidia of the modern green alga *Volvox* merits consideration. *Volvox carteri*, for example, shows a degree of spatial cellular differentiation similar to the matryoshka-stage fossils. During asexual reproduction, its gonidia undergo rapid palintomic cell cleavages to produce hollow coeloblastula-like embryos, which are embedded within an adult that consists of somatic cells[28]. At maturation, the embryos invert and are released as free-living juveniles. The somatic cells then undergo degeneration and programmed death[15]. Despite the remarkable similarity in the degree of cellular differentiation, the differences between matryoshka-stage fossils and *Volvox*
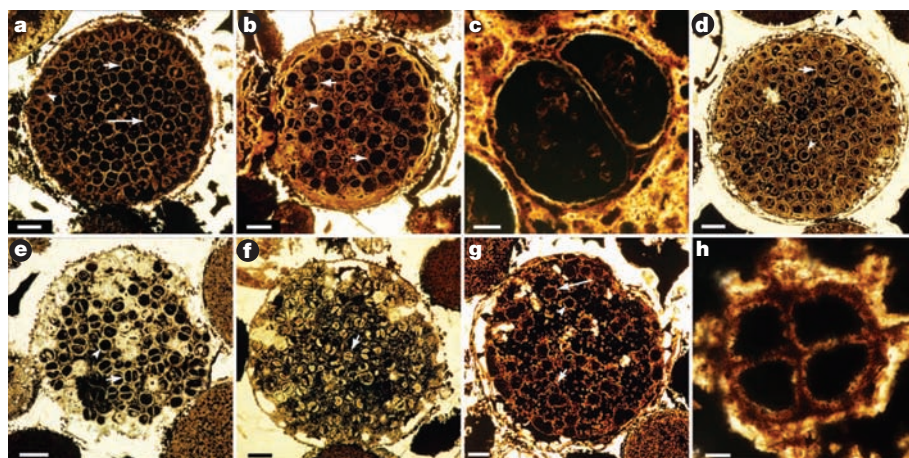
Figure 2 | *Megaclonophycus*-like fossils with dividing cell packets. a, Specimen with cell packets in the centre and slightly elongate monads in the periphery. **b–h**, Specimens with loosely packed monads, dyads, and tetrads. **c**, Magnification of dyads marked by short arrow at right in **b**. **h**, Magnification of cruciate tetrads marked by long arrow in **g**. White arrowheads, short arrows, and long arrows mark monads, dyads, and tetrads, respectively. Black arrowhead in **d** points to conical ornamentation on envelope. Scale bars: 10 μm in **c**, **h**; 100 μm in all others.
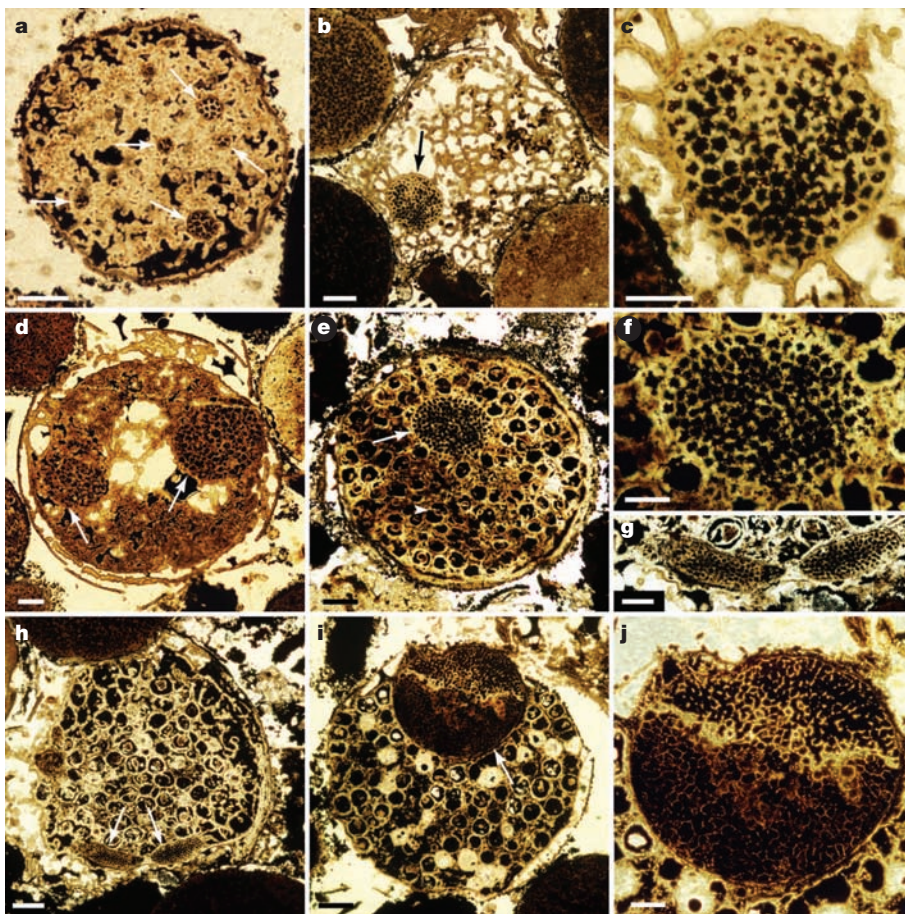
Figure 3 | *Megaclonophycus*-like fossils with matryoshkas. a, Specimen with at least five small matryoshkas. **b–j**, Specimens with one or more spheroidal or ellipsoidal matryoshkas. **c, f, g, j**, Magnifications of matryoshkas in **b, e, h, i**, respectively. Arrows mark matryoshkas, and arrowhead in **e** denotes dyads. Note nucleus-like diagenetic structures in **e, h, i**. Scale bars: 50 μm in **c, f, g, j**; 100 μm in all others.

are profound. First, gonidia of *V. cateri* undergo palintomic cell division[15], but matryoshkas presumably equivalent to gonidia are growing structures with no evidence of palintomic cell division (Extended Data Fig. 1c, d). Second, an ornamented envelope seems to persist through the *Parapandorina*-, *Megaclonophycus*-, and matryoshka-stages, implying that their cells could not have had functional locomotive flagella. In *Volvox*, an ornament envelope only exists in the dormant zygote stage during sexual reproduction, and it is shed upon meiosis and germination to facilitate growth, flagellar locomotion, and photosynthesis[15]. Third, both somatic and gonidial cells in *Volvox* are arranged peripherally, forming hollow spheres to facilitate flagellar locomotion and gonidium release[15]. In *Megasphaera*, however, both the somatic and matryoshka cells are tightly arranged to form solid multicellular structures, obstructing flagellar locomotion and impeding photosynthesis owing to self-shading. Finally, multicellular volvocines are exclusively freshwater algae[29] that diverged in the Permian-Triassic periods according to molecular clock estimates[30]. Thus, the similar degree of cellular differentiation between *Volvox* and matryoshka-stage fossils is probably convergent, and *Volvox* is a poor interpretative model for the Doushantuo fossils. However, this assessment does not exclude the possibility that the Doushantuo fossils may represent other cellularly differentiated multicellular algae.

A life cycle including a matryoshka stage excludes a phylogenetic affinity with crown-group animals, where embryogenesis does not produce a matryoshka and germ–soma separation occurs ontogenetically later in sexual reproduction. However, the present evidence does not force the falsification of the stem-group animal interpretation. Crown-group animals and their closest living sister group, the choanoflagellates, are separated by important morphological gaps. Morphological features characteristic of crown-group animals but absent in choanoflagellates—including obligate multicellularity, functional cell-to-cell adhesion, spatial cell differentiation and regionalization, germ–soma separation, apoptosis, and embryogenesis—have to evolve stepwise along the stem towards crown-group animals. The earliest stem-group animals are not expected to have all features that collectively define crown-group animals. Considering the evidence for cell-to-cell adhesion, multicellularity, spatial cell differentiation, germ–soma separation, apoptosis, and the potential lack of a rigid cell wall, it remains possible that the Doushantuo fossils could be stem-group animals that evolved an autapomorphic life cycle[24] involving a matryoshka stage.

Guided by these new fossils, our search for the phylogenetic home of the Doushantuo 'animal embryos' should focus on complex multicellular eukaryotes. Complex multicellularity evolved independently in animals, ascomycetes, basidiomycetes, and multiple green, red, and brown algal clades[26,27]. Among these groups, modern volvocine green algae and animal embryos provide partial but imperfect interpretive analogues. Future research should aim at a broader paleontological search to reconstruct the complete life cycle of these fossils and to explore other interpretive analogues of complex multicellular eukaryotes.

1. Xiao, S., Zhang, Y. & Knoll, A. H. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558 (1998).
2. Xiao, S. & Knoll, A. H. Phosphatized animal embryos from the Neoproterozoic Doushantuo Formation at Weng'an, Guizhou, South China. *J. Paleontol.* **74**, 767–788 (2000).
3. Bailey, J. V., Joye, S. B., Kalanetra, K. M., Flood, B. E. & Corsetti, F. A. Evidence of giant sulphur bacteria in Neoproterozoic phosphorites. *Nature* **445**, 198–201 (2007).
4. Bengtson, S., Cunningham, J. A., Yin, C. & Donoghue, P. C. J. A merciful death for the "earliest bilaterian," Vernanimalcula. *Evol. Dev.* **14**, 421–427 (2012).
5. Huldtgren, T. *et al.* Fossilized nuclei and germination structures identify Ediacaran "animal embryos" as encysting protists. *Science* **334**, 1696–1699 (2011).

6. Xue, Y., Tang, T., Yu, C. & Zhou, C. Large spheroidal chlorophyta fossils from the Doushantuo Formation phosphoric sequence (late Sinian), central Guizhou, South China. *Acta Palaeontol. Sin.* **34,** 688–706 (1995).
7. Butterfield, N. J. Terminal developments in Ediacaran embryology. *Science* **334,** 1655–1656 (2011).
8. Hagadorn, J. W. *et al.* Cellular and subcellular structure of Neoproterozoic embryos. *Science* **314,** 291–294 (2006).
9. Yin, L. *et al.* Doushantuo embryos preserved inside diapause egg cysts. *Nature* **446,** 661–663 (2007).
10. Cohen, P. A., Knoll, A. H. & Kodner, R. B. Large spinose microfossils in Ediacaran rocks as resting stages of early animals. *Proc. Natl Acad. Sci. USA* **106,** 6519–6524 (2009).
11. Yin, Z. *et al.* Early embryogenesis of potential bilaterian animals with polar lobe formation from the Ediacaran Weng'an Biota, South China. *Precambr. Res.* **225,** 44–57 (2013).
12. Chen, J.-Y. *et al.* Complex embryos displaying bilaterian characters from Precambrian Doushantuo phosphate deposits, Weng'an, Guizhou, China. *Proc. Natl Acad. Sci. USA* **106,** 19056–19060 (2009).
13. Xiao, S., Knoll, A. H., Yuan, X. & Pueschel, C. M. Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. *Am. J. Bot.* **91,** 214–227 (2004).
14. Xiao, S., Zhou, C., Liu, P., Wang, D. & Yuan, X. Phosphatized acanthomorphic acritarchs and related microfossils from the Ediacaran Doushantuo Formation at Weng'an (South China) and their implications for biostratigraphic correlation. *J. Paleontol.* **88,** 1–67 (2014).
15. Kirk, D. L. *Volvox: Molecular-Genetic Origins of Multicellularity and Cellular Differentiation* (Cambridge Univ. Press, 1998).
16. Marshall, W. L. & Berbee, M. L. Population-level analyses indirectly reveal cryptic sex and life history traits of *Pseudoperkinsus tapetis* (Ichthyosporea, Opisthokonta): a unicellular relative of the animals. *Mol. Biol. Evol.* **27,** 2014–2026 (2010).
17. Xiao, S. Mitotic topologies and mechanics of Neoproterozoic algae and animal embryos. *Paleobiology* **28,** 244–250 (2002).
18. Xiao, S., Knoll, A. H., Schiffbauer, J. D., Zhou, C. & Yuan, X. Comment on ''Fossilized nuclei and germination structures identify Ediacaran 'animal embryos' as encysting protists''. *Science* **335,** 1169c (2012).
19. Xiao, S., Yuan, X. & Knoll, A. H. Eumetazoan fossils in terminal Proterozoic phosphorites? *Proc. Natl Acad. Sci. USA* **97,** 13684–13689 (2000).
20. Bengtson, S. & Budd, G. Comment on ''Small Bilaterian Fossils from 40 to 55 million Years Before the Cambrian''. *Science* **306,** 1290a–1291a (2004).
21. Cunningham, J. A. *et al.* Distinguishing geology from biology in the Ediacaran Doushantuo biota relaxes constraints on the timing of the origin of bilaterians. *Proc. R. Soc. B* **279,** 2369–2376 (2012).
22. Schiffbauer, J. D., Xiao, S., Sen Sharma, K. & Wang, G. The origin of intracellular structures in Ediacaran metazoan embryos. *Geology* **40,** 223–226 (2012).
23. Raff, E. C., Vilinski, J. T., Turner, F. R., Donoghue, P. C. J. & Raff, R. A. Experimental taphonomy shows the feasibility of fossil embryos. *Proc. Natl Acad. Sci. USA* **103,** 5846–5851 (2006).
24. Mikhailov, K. V. *et al.* The origin of Metazoa: a transition from temporal to spatial cell differentiation. *Bioessays* **31,** 758–768 (2009).
25. Boyle, R. A., Lenton, T. M. & Williams, H. T. P. Neoproterozoic 'snowball Earth' glaciations and the evolution of altruism. *Geobiology* **5,** 337–349 (2007).
26. Knoll, A. H. The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* **39,** 217–239 (2011).
27. Herron, M. D., Rashidi, A., Shelton, D. E. & Driscoll, W. W. Cellular differentiation and individuality in the 'minor' multicellular taxa. *Biol. Rev. Camb. Philos. Soc.* **88,** 844–861 (2013).
28. Green, K. J. & Kirk, D. L. Cleavage patterns, cell lineages, and development of a cytoplasmic bridge system in *Volvox* embryos. *J. Cell Biol.* **91,** 743–755 (1981).
29. Umen, J. G. & Olsony, B. J. S. C. Genomics of volvocine algae. *Adv. Bot. Res.* **64,** 185–243 (2012).
30. Herron, M. D., Hackett, J. D., Aylward, F. O. & Michod, R. E. Triassic origin and early radiation of multicellular volvocine algae. *Proc. Natl Acad. Sci. USA* **106,** 3254–3258 (2009).

Author Contributions L.C. and K.P. performed the research under the guidance of S.X. and X.Y. S.X. developed the interpretation and prepared the manuscript with the assistance of C.Z. and X.Y.

METHODS

Fossils were collected from black phosphorite (unit 4A of ref. 14) of the Doushantuo Formation at Weng'an, a unit overlooked in previous studies, which were mostly limited to specimens extracted from the grey phosphorite (unit 4B of ref. 14). Because of silica and phosphate cementation, phosphatized fossils in the black phosphorite cannot be extracted using acid extraction techniques. Thus, fossils were examined in thin sections and their three-dimensional morphologies were inferred from the observation of multiple specimens at random orientations. Fossils in thin sections were examined and photographed on a Zeiss Axioskop 2 plus microscope attached with a digital camera. No digital manipulations other than adjustment of brightness and contrast have been applied to the photographs.

31. Xiao, S., Zhou, C. & Yuan, X. Undressing and redressing Ediacaran embryos. *Nature* **446,** E9–E10 (2007).
32. Liu, P., Yin, C., Chen, S., Tang, F. & Gao, L. New data of phosphatized globular fossils from Weng'an biota in the Ediacaran Doushantuo Formation and the problem concerning their affinity. *Acta Geosci. Sin.* **30,** 457–464 (2009).
33. Yin, Z. & Zhu, M. Quantitative analysis on the fossil abundance of the Ediacaran Weng'an biota, Guizhou. *Acta Palaeontol. Sin.* **47,** 477–487 (2008).
34. Zhou, C., Yuan, X. & Xiao, S. Phosphatized biotas from the Neoproterozoic Doushantuo Formation on the Yangtze Platform. *Chin. Sci. Bull.* **47,** 1918–1924 (2002).
35. Yin, Z., Liu, P., Li, G., Tafforeau, P. & Zhu, M. Biological and taphonomic implications of Ediacaran fossil embryos undergoing cytokinesis. *Gondwana Res.* **25,** 1019–1026 (2014).
36. Li, S., Zhao, J., Lu, P. & Xie, Y. Maximum packing densities of basic 3D objects. *Chin. Sci. Bull.* **55,** 114–119 (2010).
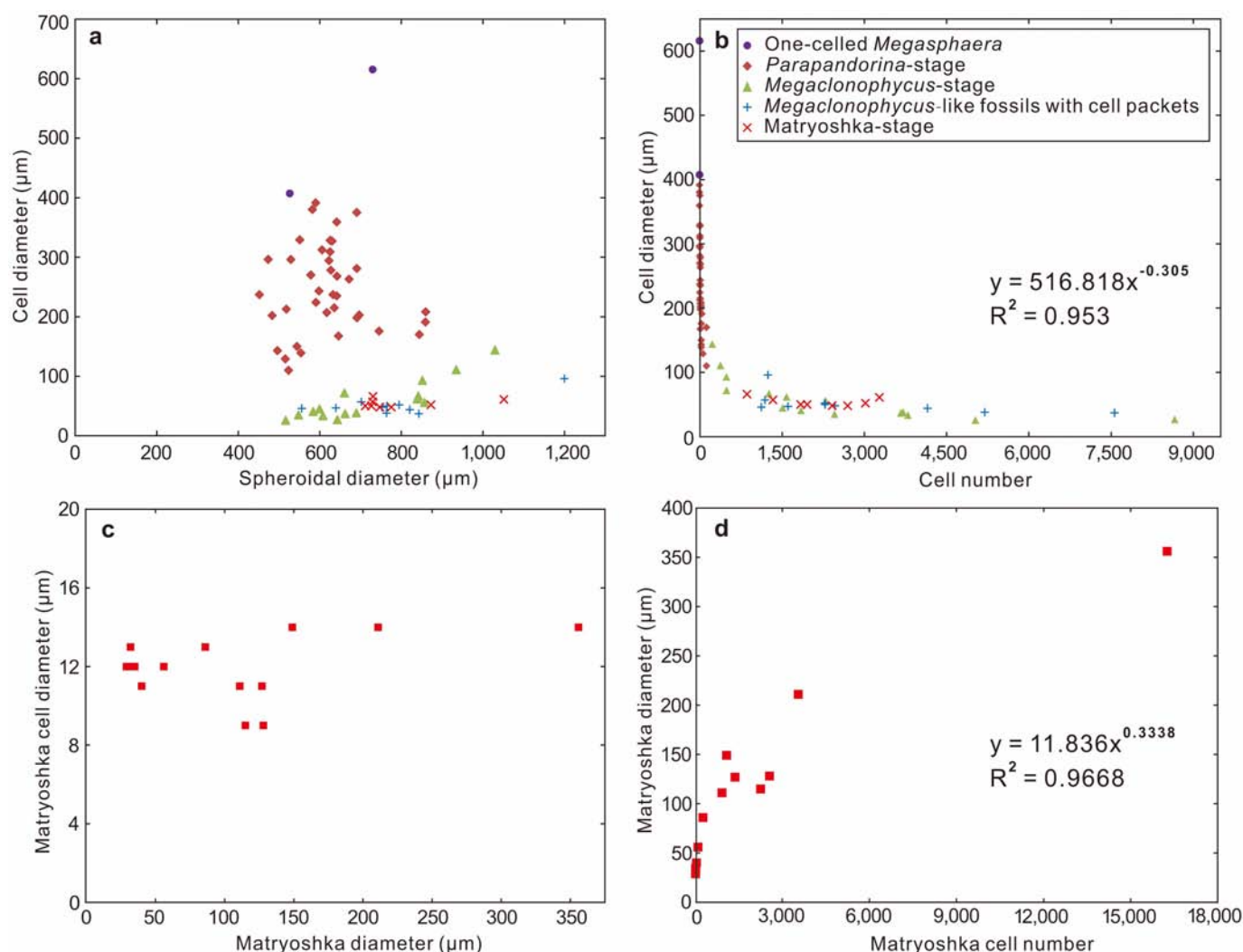
Extended Data Figure 1 | Measurements of *Megaclonophycus*-like fossils with cell packets and matryoshkas. a, b, Cross-plots of specimen size (that is, diameter of spheroidal fossils), cell number, and diameter of blastomere-like cells, showing the constancy of spheroidal size, independency of spheroidal size on cell size, and power relationship between cell number and cell diameter, as predicted from palintomic cell division. The relationship confirms that the *Megaclonophycus*-like fossils with cell packets and matryoshkas follow an ontogenetic trajectory established by *Parapandorina*- and *Megaclonophycus*-stage fossils. Measurements were made on thin-section specimens in our collection as well as on extracted specimens from published material[1,2,31–35]. Each data point represents a single specimen, with its diameter averaged between maximum and minimum dimensions. Cell diameter is averaged among all observable cells, excluding cell packets and matryoshkas. In *Parapandorina*-stage specimens, cell number was determined from actual count whenever possible. In *Megaclonophycus*-stage specimens, cell number

was estimated from random close packing of spherical cells (64% packing density[36]). In *Megaclonophycus*-like specimens with cell packets and matryoshkas, cell number was also estimated from random close packing of spherical cells, but assuming that the volumes of cell packets or matryoshkas were occupied by spherical blastomere-like cells. **c, d**, Cross-plots of matryoshka diameter, cell number in matryoshkas, and average cell size in matryoshkas, showing constancy of cell size, independency of matryoshka size on cell size, and power relationship between matryoshka diameter and cell number, as predicted from the continuing growth of matryoshkas. Each data point represents a single matryoshka, with its diameter averaged between its maximum and minimum dimensions. Cell diameter is averaged among all observable cells encountered in thin sections. Cell number was estimated from tight packing of polyhedral cells (100% packing density). See Source Data for measurements.

LETTER

An evolutionary arms race between KRAB zinc–finger genes *ZNF91/93* and SVA/L1 retrotransposons

Frank M. J. Jacobs[1]*†, David Greenberg[1,2]*†, Ngan Nguyen[1,3], Maximilian Haeussler[1], Adam D. Ewing[1]†, Sol Katzman[1], Benedict Paten[1], Sofie R. Salama[1,4] & David Haussler[1,4]

Throughout evolution primate genomes have been modified by waves of retrotransposon insertions[1–3]. For each wave, the host eventually finds a way to repress retrotransposon transcription and prevent further insertions. In mouse embryonic stem cells, transcriptional silencing of retrotransposons requires KAP1 (also known as TRIM28) and its repressive complex, which can be recruited to target sites by KRAB zinc–finger (KZNF) proteins such as murine-specific ZFP809 which binds to integrated murine leukaemia virus DNA elements and recruits KAP1 to repress them[4,5]. KZNF genes are one of the fastest growing gene families in primates and this expansion is hypothesized to enable primates to respond to newly emerged retrotransposons[6,7]. However, the identity of KZNF genes battling retrotransposons currently active in the human genome, such as SINE-VNTR-Alu (SVA)[8] and long interspersed nuclear element 1 (L1)[9], is unknown. Here we show that two primate-specific KZNF genes rapidly evolved to repress these two distinct retrotransposon families shortly after they began to spread in our ancestral genome. ZNF91 underwent a series of structural changes 8–12 million years ago that enabled it to repress SVA elements. ZNF93 evolved earlier to repress the primate L1 lineage until ~12.5 million years ago when the L1PA3-subfamily of retrotransposons escaped ZNF93's restriction through the removal of the ZNF93-binding site. Our data support a model where KZNF gene expansion limits the activity of newly emerged retrotransposon classes, and this is followed by mutations in these retrotransposons to evade repression, a cycle of events that could explain the rapid expansion of lineage-specific KZNF genes.

KAP1 mediates transcriptional silencing of retrotransposons and protects genome integrity through repression of retrotransposition activity[10,11]. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis revealed that in human embryonic stem cells (hESCs), KAP1 predominantly associates with active primate-specific classes of retrotransposons such as SVA and L1PA (Extended Data Fig. 1)[11,12]. Similarly, in mouse ESCs (mESCs) KAP1 primarily associates with mouse-lineage-specific retrotransposon classes (Extended Data Fig. 2)[12]. These data support the hypothesis that species-specific KZNFs recruit KAP1 to species-specific retrotransposon classes that recently invaded the host's genome[7,13]. To test this, we determined the fate of primate-specific retrotransposons in a non-primate background using trans-chromosomic mESCs that contain a copy of human chromosome 11 (E14(hChr11) cells)[14], hereafter termed trans-chromosomic 11 (TC11)-mESCs. In the TC11-mESC cellular environment, primate-specific retrotransposons, including SVA and L1PA elements, are derepressed and gain activating histone H3 Lys 4 (H3K4me3) marks (Fig. 1a, b and Extended Data Fig. 1e). As a result of this de-repression, a majority of SVA (51%), human-specific L1 (L1Hs) (93%) and some other L1PA elements, such as L1PA4 (16%), become aberrantly transcribed. These findings suggest primate-specific retrotransposons have a transcriptional potential[15,16] that is repressed by primate-specific factors.

Promising candidates for these factors are the approximately 170 KZNF genes that emerged during primate evolution[7] (Extended Data Fig. 3a). We reasoned that a KZNF gene responsible for protecting genome integrity, most critical in the germ line, must be highly expressed in hESCs. So we focused on 14 highly expressed, primate-specific KZNF genes (Extended Data Fig. 3b) and tested each candidate for a role in repressing SVA retrotransposons, which first appeared in great apes 18–25 million years (Myr) ago[8], and are still active[17]. We set up a luciferase assay based screen in mESCs in which an SVA element cloned upstream of a minimal SV40 promoter strongly enhances luciferase activity (Extended Data Fig. 4a). Each candidate KZNF was co-expressed with the SVA–luciferase construct to determine its effect on reporter activity. Of all KZNFs tested, ZNF91 most dramatically decreased SVA-driven luciferase activity, reducing activity to $16 \pm 4\%$ relative to an empty-vector-transfected control (Fig. 2a). Some other KZNFs had modest effects on this reporter, but were not further analysed, as those with the strongest effect also inhibited the OCT4 (also known as POU5F1) enhancer, which is not KAP1-bound in ESCs, and therefore suggests a nonspecific effect (Extended Data Fig. 7a). Structure–function analysis of SVA revealed that the variable number tandem repeat (VNTR) domain is necessary
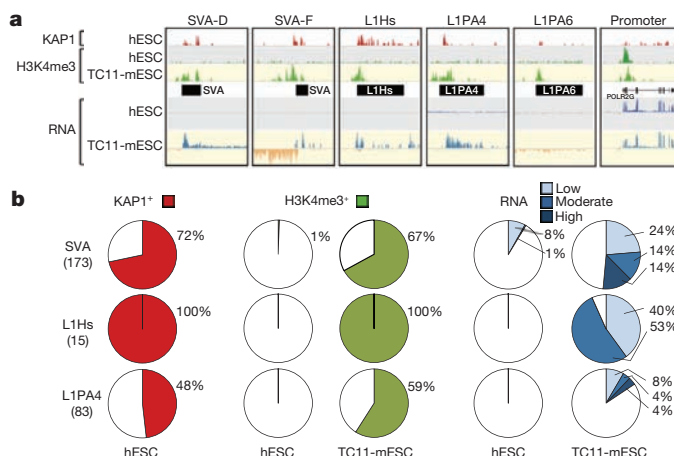

Figure 1 | SVAs and L1PAs are derepressed in a non-primate cellular environment. a, KAP1, H3K4me3 ChIP-seq and RNA sequencing (RNA-seq) coverage tracks for a selection of KAP1-bound primate-specific retrotransposons derepressed in TC11-mESCs (yellow) relative to hESCs (grey). H3K4me3 signal on promoters is similar in hESCs and TC11-mESCs. **b**, Percentages of SVA, L1Hs and L1PA elements on human chromosome 11 positive for KAP1, H3K4me3 and relative levels of transcription (see Methods) in hESC and TC11-mESCs. Total elements of each type on human chromosome 11 in parentheses.

[1]Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [2]Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, California 95064, USA. [3]Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [4]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. †Present addresses: Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam 1098 XH, The Netherlands (F.M.J.J.); Gladstone Institute of Virology and Immunology, San Francisco, California 94158, USA (D.G.); Mater Research Institute, University of Queensland, Queensland 4101, Australia (A.D.E.).
*These authors contributed equally to this work.
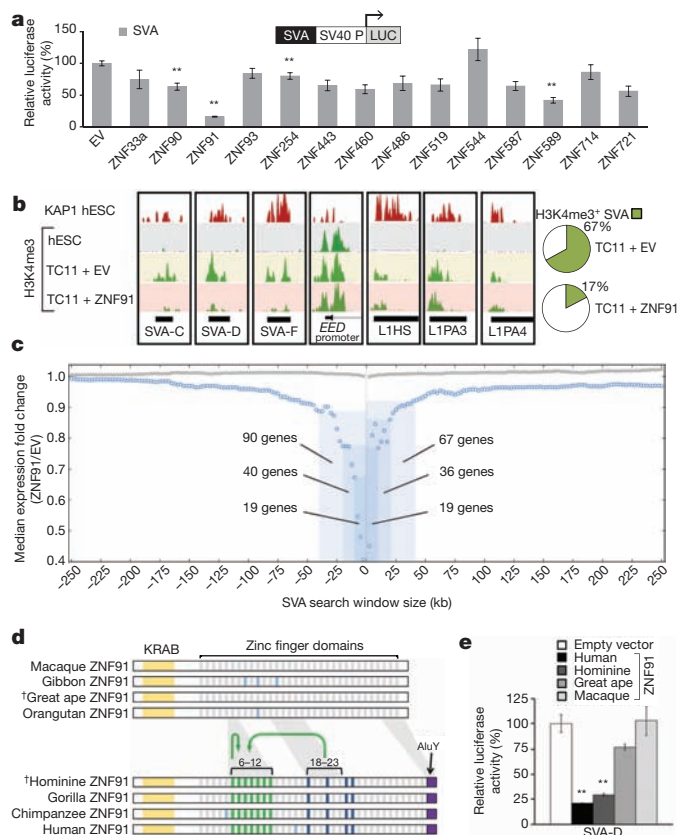
Figure 2 | SVA elements are repressed by primate-specific ZNF91.
a, Relative luciferase activity of a SVA-D–SV40–luciferase-reporter after co-transfection of KZNFs in mESCs. EV, empty vector. **b**, KAP1 and H3K4me3 ChIP-seq coverage tracks for a selection of loci in hESCs and TC11-mESCs transfected with an empty vector (TC11 + EV) or ZNF91 (TC11 + ZNF91). Pie charts show percentages of H3K4me3-positive SVAs on human chromosome 11. **c**, Median fold expression change (ZNF91 relative to empty vector), for genes with (blue circles) or without (grey crosses) an SVA within the indicated genomic distance among the 994 expressed human chromosome 11 genes; kb, kilobases. **d**, ZNF91 structural evolution. Green stripes, duplicated zinc-fingers; blue stripes, zinc-fingers that changed contact residues in the lineage to humans (dark blue) or in other lineages (light blue). Green arrows indicate segmental duplications. Dagger symbols indicate reconstructed ancestral proteins. **e**, Relative SVA_D–SV40–luciferase activity in the presence of various ZNF91 proteins. **a**, **e**, ******$P < 0.01$; error bars are s.e.m.

and sufficient for ZNF91-mediated repression of luciferase activity (Extended Data Fig. 4b, c). Furthermore, transfection of TC11-mESCs with human ZNF91 restored the repression of deregulated SVAs on human chromosome 11, causing a strong decrease of the aberrant H3K4me3 ChIP-seq signal at SVAs, while leaving other derepressed elements such as L1Hs or L1PAs unaffected (Fig. 2b and Extended Data Fig. 5a). Transfection of ZNF91 also significantly repressed aberrant transcription of SVA repeats, indicating that ZNF91 is sufficient to restore transcriptional silencing of SVAs. (Extended Data Fig. 5b). No such effects were observed for other primate KZNFs (*ZNF90*, *ZNF93*, *ZNF486*, *ZNF826*, *ZNF443*, *ZNF544* or *ZNF519*) transfected in TC11-mESCs, validating the specificity of the ZNF91–SVA interaction (Extended Data Fig. 5c). Cellular genes near SVAs on human chromosome 11 in TC11-mESCs were also repressed by ZFN91, with the distance of a gene to an SVA as the major factor governing the amount of bystander repression (Fig. 2c), supporting the hypothesis that the host response to retrotransposon insertion has significantly impacted human gene expression patterns[11,15,16].

ZNF91 emerged in the last common ancestor (LCA) of humans and Old-World monkeys and has undergone dramatic structural changes, including the addition of seven zinc-fingers in the LCA of humans and gorillas[18] (Fig. 2d). We reconstructed ancestral versions of ZNF91 by

parsimony analysis (Extended Data Fig. 6a, b) and found that ZNF91 as it probably existed in the LCA of humans and gorillas (ZNF91[hominine]) was able to repress the SVA-luciferase reporter in a similar fashion to human ZNF91 (Fig. 2e). However, ZNF91 as it existed in the LCA of humans and orangutans (ZNF91[great ape]) only reduced luciferase activity to around 80% of baseline and macaque ZNF91 completely lacked the ability to repress SVA-driven luciferase activity. The importance of the seven recently added hominine zinc-fingers was further supported by deletion analysis of ZNF91 (Extended Data Fig. 6c). These findings suggest that the changes in ZNF91 between 8–12 Myr ago have markedly improved the protein's ability to bind and repress SVA.

In our KAP1 ChIP experiments, KAP1 also showed a strong association with the 5′ untranslated region (UTR) of L1PA elements. None of the 14 KZNFs had a significant effect on the 5′ UTR of the current active L1Hs[9,19] cloned upstream of the luciferase reporter when tested in mESCs. However, ZNF93 significantly reduced luciferase activity of a reporter with the 5′ UTR of a KAP1-positive L1PA4 element (62 ± 10%, Extended Data Fig. 7a). To verify the recruitment of ZNF93 to L1PA4 elements on the human genome, we performed ChIP-seq analysis on hESCs using antibody ab104878, which recognizes ZNF93 and co-immunoprecipitates KAP1 (Extended Data Fig. 7b, c). We found that ZNF93 binds to the 5′ end of L1PA4, the ancestral subtypes L1PA6 and L1PA5, and the descendant subtype L1PA3 (Fig. 3a and Extended Data Fig. 7d). To validate that the ab104878 ChIP-seq signal on L1PAs is derived from ZNF93, we performed ab104878-ChIP analysis followed by quantitative PCR on TC11-mESC transfected with ZNF93 or an empty vector and found significant enrichment of the L1PA4 5′ UTR compared to a LTR12C control element (Extended Data Fig. 7e). No consistent ZNF93 binding was detected at L1PA7 or older subtypes nor at the most recently evolved L1PA2 and L1Hs (Fig. 3a). Comparative sequence analysis revealed that the absence of ZNF93 binding in L1Hs and L1PA2 can be explained by a 129-base-pair (bp) deletion in the 5′ UTR that spans the ChIP-determined ZNF93- and KAP1-binding sites (Fig. 3b). The deletion is also present in ~50% of L1PA3 elements, resulting in distinct subgroups of shorter (L1PA3-6030) and longer (L1PA3-6160) L1PA3 elements, but is not present in L1PA4–6 families.

To investigate the interaction of ZNF93 with the 129-bp L1PA element, we tested a series of L1PA4 segments cloned upstream of an OCT4-enhancer fused to an SV40-promoter and luciferase-reporter in mESCs (Fig. 3c). Both the 129-bp element and a 51-bp sub-fragment were sufficient to confer ZNF93-mediated repression of the luciferase reporter, and this repression was abolished by elimination of the 51-bp portion in the 129-bp fragment (129Δ51[L1PA4]). The 51-bp element encompasses a computationally predicted DNA binding motif for the 17 fingers of ZNF93[20] and the central 18 bp of this region displays strong similarity to the predicted recognition motif of zinc-fingers 8–13 of human ZNF93 (Fig. 3d). A ZNF93 variant that has all contact residues in zinc-fingers 8–13 replaced by serine residues (ZNF93serF), a modification that abolishes DNA binding selectivity[21], was unable to repress luciferase activity of the L1PA4 elements (Fig. 3e), suggesting that fingers 8–13 of ZNF93 are important for recognition of the 129-bp element in L1PA3-6 retrotransposons.

ZNF93 emerged in the LCA of apes and Old-World monkeys and reconstruction of the evolutionary history of the ZNF93 protein by parsimony suggests that dramatic changes took place in the LCA of orangutans and humans between 12–18 Myr ago (ZNF93[great ape]; Extended Data Fig. 8a). Indeed, macaque ZNF93 does not have the ability to repress the 129-bp or 51-bp element of L1PA4 in the luciferase assay, but ZNF93[great ape] represses at levels similar to ZNF93[human] (Extended Data Fig. 8b), suggesting changes in the ape lineage probably enabled ZNF93 to regulate L1 activity.

To explore the function of the lost 129-bp element, we created a version of L1Hs with this sequence restored in its 5′ UTR (L1Hs+129[L1PA4]), or a scrambled version of this 129-bp sequence (L1Hs+129scramble[L1PA4]) as a control, and compared retrotransposition efficiencies to wild-type L1Hs in HEK293FT cells in an *in vitro* retrotransposition assay[22,23]. In
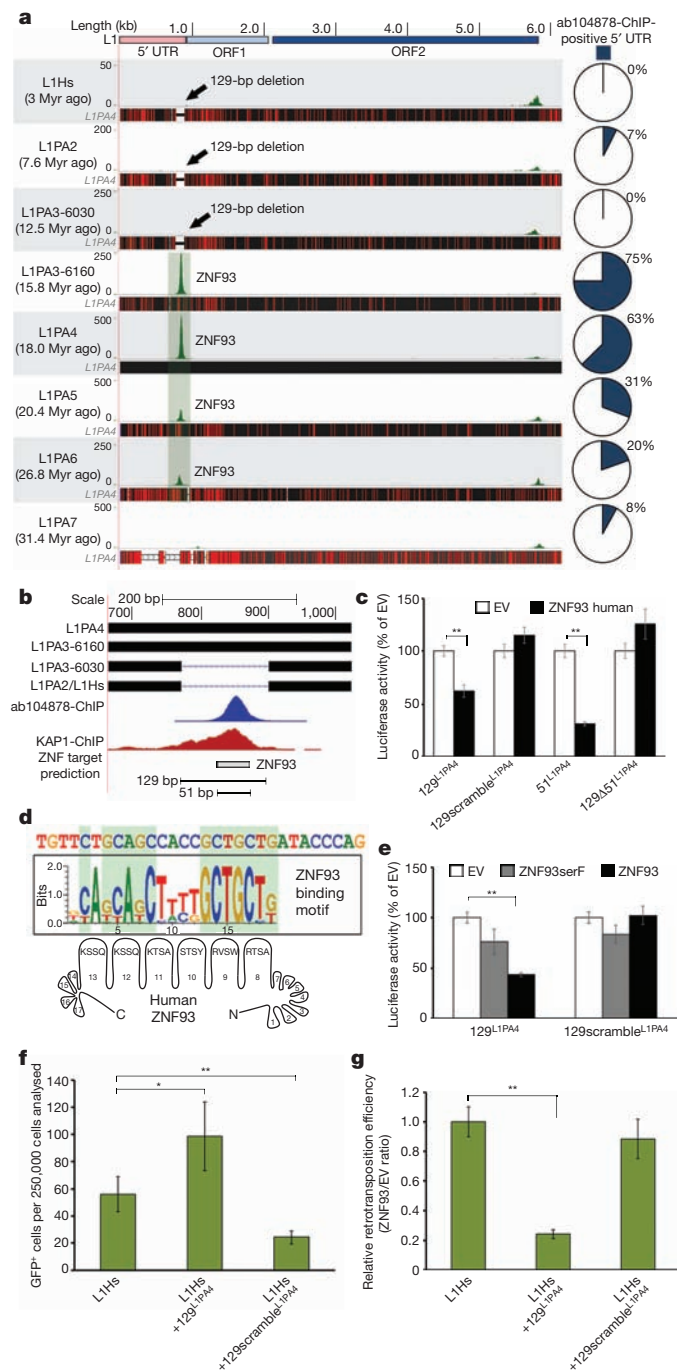
Figure 3 | L1PA elements are repressed by primate-specific ZNF93. a, Green peaks represent genome-wide ab104878-ChIP-seq peak-summits mapped to L1PA consensus sequences. Black horizontal bars, alignment to L1PA4; red lines, divergent positions. **b,** The 129-bp deletion and predicted 51-bp ZNF93 binding motif (grey bar) relative to L1PA4. **c,** Relative activity of OCT4-enhancer–luciferase-reporters after co-transfection of an empty vector (EV) or ZNF93. 129^{L1PA4}, 129-bp fragment of L1PA4; $129\Delta51^{L1PA4}$, 129-bp fragment without the 51-bp part; $129scramble^{L1PA4}$, scrambled 129-bp fragment; 51^{L1PA4}, 51-bp fragment. **d,** Consensus central sequence of ab104878-ChIP-seq summits for L1PA4, aligned with the predicted recognition motif of ZNF93 zinc-fingers 8–13. **e,** Relative activity for OCT4-enhancer–luciferase-reporters after co-transfection of EV, ZNF93serF or ZNF93. **f,** Number of GFP-positive cells derived from retrotransposition events of L1Hs, L1Hs+129 and L1Hs+129scrambled constructs in HEK cells ($n = 7$). **g,** Same as **f** but showing the ratio of retrotransposition events after co-transfection with ZNF93 compared to an empty vector. **c, e, f, g,** $*P < 0.05$; $**P < 0.01$; error bars are s.e.m.

this assay, a retrotransposition event results in green fluorescent protein (GFP) expression (Extended Data Fig. 9). $L1Hs+129^{L1PA4}$ shows a 1.76-fold (± 0.45 s.e.m.) higher retrotransposition activity compared to wild-type L1Hs, an effect not seen with $L1Hs+129scramble^{L1PA4}$ (Fig. 3f), suggesting that this 129-bp sequence promotes retrotransposition. Importantly, co-expression of ZNF93 significantly reduced retrotransposition of $L1Hs+129^{L1PA4}$ to just 24% ($\pm 3\%$ s.e.m.) relative to L1Hs, but had no significant effect on $L1Hs+129scramble^{L1PA4}$ (Fig. 3g).

These data suggest the 129-bp sequence, as it once existed in the 5′ UTR of L1PA subfamilies, may have been beneficial to L1 mobilization, but since ZNF93 evolved to bind this element, losing it allowed the L1 lineage to escape ZNF93-mediated repression, providing net selective advantage. Indeed, phylogenetic analysis of L1PA3 elements and calculation of the average distance of L1PA3-6030 and L1PA3-6160 elements from the respective consensus sequences, suggests that L1PA3-6030 elements lacking the 129-bp element have expanded more recently in our genome than L1PA3-6160 elements, showing an estimated age of 12.5 and 15.8 million years, respectively (Extended Data Fig. 10a). This strongly suggests that loss of the ZNF93-binding site—and thereby the evasion of the host repression—propagated a new wave of L1 insertions in great ape genomes.
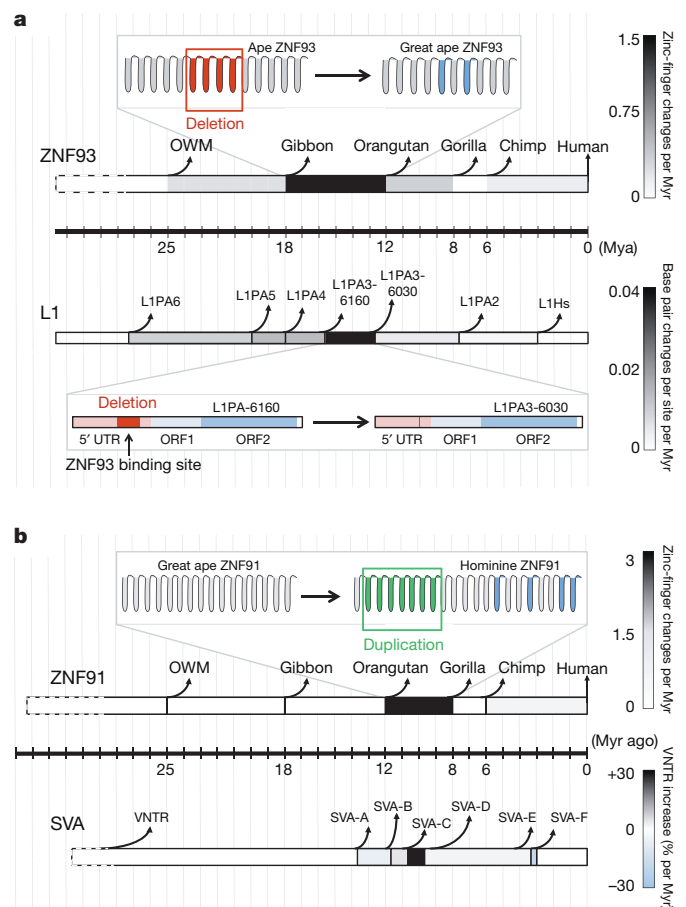
Figure 4 | Dynamic patterns of co-evolution between ZNFs and target retrotransposons. a, b, Schematic showing the evolution of L1PA[9] and SVA[8] retrotransposons parallel to the structural evolution of ZNF93 and ZNF91 along an evolutionary timescale. Colouring of ZNF91 and ZNF93 horizontal bars represent zinc-finger changes per million years during the time interval indicated. Red zinc-fingers, deletion; blue zinc-fingers, change in contact residues; green zinc-fingers, duplication. Colouring of retrotransposon horizontal bars represents base-pair substitutions, deletions or insertions per site per million years (L1PA), or percentage increase in VNTR size per million years (SVA). Myr, million years; OWM, Old-World monkey.

Repeated turnover of the 5′ UTR occurred in early L1PA evolution[9] and was previously thought to be associated with competition for host factors[24]. Our results suggest turnover was instead driven by avoidance of host factors. The precise removal of the ZNF93-binding site probably took place soon after ZNF93 underwent a series of structural changes, suggesting the deletion may have been driven by improved host repression of L1PA activity (Fig. 4a). In a similar fashion, the structural changes in ZNF91 allowing it to repress SVA elements may have driven the further evolution of new and different SVA-subtypes in gorillas, chimpanzees and humans, a pattern that is not observed in orangutans, which diverged before ZNF91 had undergone these structural changes (Extended Data Fig. 10b). Notably, the size of the VNTR region of SVA, the prime interaction site of ZNF91, has increased during the timeframe of structural changes to ZNF91 (Fig. 4b and Extended Data Fig. 10c).

Our data support a model in which modifications to lineage-specific KZNF genes are used by the host to repress new families of retrotransposons as they emerge, which in turn drives the evolution of newer families of retrotransposons, in a continuing arms race. Because repression affects nearby genes, KZNFs have probably been co-opted for other functions that persisted long after the original transposon expansion they first evolved to repress had subsided[25], fuelling the evolution of more complex gene-regulatory networks. Unlike an arms race with an external pathogen, retrotransposons are host DNA, suggesting that a mammalian genome is itself in an internal arms race with its own DNA, and thereby inexorably driven towards greater complexity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Kazazian, H. H. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
2. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10,** 691–703 (2009).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
4. Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131,** 46–57 (2007).
5. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458,** 1201–1204 (2009).
6. Birtle, Z. & Ponting, C. P. Meisetz and the birth of the KRAB motif. *Bioinformatics* **22,** 2841–2845 (2006).
7. Thomas, J. H. & Schneider, S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* **21,** 1800–1812 (2011).
8. Wang, H. *et al.* SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* **354,** 994–1007 (2005).
9. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16,** 78–87 (2006).
10. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463,** 237–240 (2010).
11. Turelli, P. *et al.* Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **24,** 1260–1270 (2014).
12. Castro-Diaz, N. *et al.* Evolutionarily dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28,** 1397–1409 (2014).
13. Huntley, S. *et al.* A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* **16,** 669–677 (2006).
14. Kai, Y. *et al.* Enhanced apoptosis during early neuronal differentiation in mouse ES cells with autosomal imbalance. *Cell Res.* **19,** 247–258 (2009).
15. Gifford, W. D., Pfaff, S. L. & Macfarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23,** 218–226 (2013).
16. Ward, M. C. *et al.* Latent regulatory potential of human-specific repetitive elements. *Mol. Cell* **49,** 262–272 (2013).
17. Hancks, D. C. & Kazazian, H. H. Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22,** 191–203 (2012).
18. Bellefroid, E. J. *et al.* Emergence of the ZNF91 Krüppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. *Proc. Natl Acad. Sci. USA* **92,** 10757–10761 (1995).
19. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nature Rev. Genet.* **12,** 615–627 (2011).
20. Persikov, A. V., Osada, R. & Singh, M. Predicting DNA recognition by Cys_2His_2 zinc finger proteins. *Bioinformatics* **25,** 22–29 (2009).
21. Moore, M., Choo, Y. & Klug, A. Design of polyzinc finger peptides with structured linkers. *Proc. Natl Acad. Sci. USA* **98,** 1432–1436 (2001).
22. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28,** 1418–1423 (2000).
23. Kimberland, M. L. *et al.* Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum. Mol. Genet.* **8,** 1557–1560 (1999).
24. Swergold, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10,** 6718–6729 (1990).
25. Lowe, C. B., Bejerano, G. & Haussler, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl Acad. Sci. USA* **104,** 8005–8010 (2007).

METHODS

Embryonic stem cell culture and ZNF overexpression analysis. Human (H9) ESC colonies were maintained as described (http://www.wicell.org). Colonies were manually passaged at a 1:3 ratio onto plates containing mitomycin-C-treated mouse embryonic fibroblasts that were seeded at a density of 35,000 cells cm^{-2} on 0.25% gelatin-coated plates (porcine; Sigma) the day before. Mouse transchromosomic E14(hChr11) (TC11) ESCs were cultured on mouse embryonic fibroblast feeder layers as described[14]. For transfections, cells were cultured on gelatin for two passages and transfected with 24 μg of ZNF and 1 μg of GFP expression vectors per 10 cm plate of cells, using lipofectamine 2000 (Invitrogen). Cells were cultured for an additional 40 h, harvested with trypleE reagent (Life technologies) and washed three times and collected in fluorescence-activated cell sorting (FACS) buffer (1× PBS, 2% fetal bovine serum (FBS), 5 mM EDTA). GFP-positive cells were sorted using a FACSAria III (BD Biosciences) and samples were used for RNA isolation and ChIP analysis.

RNA-seq library preparation. RNA was treated with RQ1 DNaseI (Promega) for 1 h at 37 °C and total RNA was cleaned up using the RNAeasy Mini kit (Qiagen). For each sample, the non-ribosomal fraction of 5 μg of total RNA was isolated using a Ribo-Zero rRNA removal Kit (Epicentre) following the manufacturer's protocol (Lit. 309-6/2011). For the non-ribosomal fraction of RNA, double stranded (ds) complementary DNA was synthesized as described previously[26] using dUTP in the second strand synthesis and USER digest before amplification to retain strand specificity. Clean-up steps were performed using RNA Clean & Concentrator or DNA Clean & Concentrator kits (Zymo Research). Double stranded cDNA was used for library preparation following the Low Throughput Guidelines of the TruSeq DNA Sample Preparation kit (Illumina), with the following additions. Size selections were performed before and after cDNA amplification on an E-gel Safe Imager (Invitrogen) using 2% E-gel SizeSelect gels (Invitrogen). The cDNA fraction of 300–400 bp in size (including adapters) was isolated and purified. For adaptor ligations, 1 μl instead of 2.5 μl of DNA Adaptor Index was used. Indexed libraries were pooled and sequenced on the Illumina HiSEQ platform. Two biological replicate samples were analysed for empty-vector-transfected cells and ZNF91-transfected cells, three biological replicate samples were analysed for human ESCs and two for rhesus macaque LYON-ES1 ESCs. Data can be viewed on the UCSC browser: http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=http://hgwdev.soe.ucsc.edu/~max/jacobs 2014/hub.txt&position=chr11:60180780-60680779.

Mapping and analysis of RNA-seq data. All samples were mapped using Tophat2 (ref. 27) with Bowtie2 (ref. 28) as the underlying alignment tool. The input Illumina fastq files consisted of paired-end reads with each end containing 100 bp. The target genome assembly for the human samples was GRCh37/UCSC-hg19 for hESCs, or a hybrid target genome of mm9-hChr11 for TC11-mESCs, and Tophat was additionally supplied with a gene model (using its '-GTF' parameter) with data from the hg19 UCSC KnownGenes track[29]. For multiply-mapped fragments, only the highest scoring mapping determined by Bowtie2 was kept. Only mappings with both read ends aligned were kept. Potential PCR duplicates (mappings of more than one fragment with identical positions for both read ends) were removed with the samtools 'rmdup'[30] function, keeping only one of any potential duplicates. The final set of mapped paired-end reads for a sample were converted to position-by-position coverage of the relevant genome assembly using the bedtools 'genomeCoverageBed'[31] function. To determine the count of fragments mapping to a gene, the position-by-position coverage was summed over the exonic positions of the gene. This gene total coverage was divided by a factor of 200, to account for the 200 bp of coverage induced by each mapped paired-end fragment (100 bp from each end), and rounded to an integer. For the human samples, this was calculated for each gene in the UCSC Known Gene set. For input to DESeq[32] all genes with non-zero counts in any sample were considered. Two replicates of each sample were combined per the DESeq methodology.

For Fig. 2c, the median fold change in expression (ZNF91/EV, vertical axis) for genes with an SVA element within some distance (blue circles) and genes without an SVA element within the same distance (grey crosses) were plotted against the up- or downstream distance from each gene. A total of 994 expressed genes were considered. Points were computed every 2.5 kb, For every window size starting at 2.5 kb and progressing cumulatively up to 250 kb in 2.5 kb intervals upstream and downstream of genes on chromosome 11, we identified the set of genes with and without at least one SVA element within the window. For the two sets (genes with SVA and genes without SVA), at every window size we calculated the median fold change in gene expression (ZNF91/EV) using the DESeq results from TC11-mESCs transfected with either ZNF91 or an empty vector. The python script to generate the figure and the associated data are available at http://hgwdev.sdsc.edu/~ewingad/Tc11SVAFig2e.tar.gz.

Chromatin immunoprecipitation (ChIP), ChIP-qPCR and ChIP-seq library preparation. Human (H9) and mouse ESCs (46C and transchromosomic TC11) were crosslinked in 1% formaldehyde for 10 min on ice by adding 1/10 volume of freshly prepared 11× crosslinking solution (50 mM Hepes (pH 8.0); 0.1 M NaCl; 1 mM EDTA; 0.5 mM EGTA; 11% formaldehyde). The crosslinking reaction was quenched by adding glycine to a final concentration of 0.125 M and incubating for 5 min on ice. For KAP1-ChIP and ChIP with the KZNF antibody ab104878, cells were washed three times in PBS + 0.1% BSA and dissolved in ten packed cell volumes 0.3% SDS-lysis buffer (10 mM Tris (pH 8.0); 1 mM EDTA (pH 8.0); 0.3% (w/v) SDS + Complete Proteinase Inhibitor Cocktail (Roche)). Cells were incubated on ice for 20 min and cells were lysed in a pre-chilled Dounce homogenizer by ten strokes with pestle B. Cell lysate was transferred to a 15 ml conical (hESC) or 1.5 ml tube (mESC) and chromatin was sheared to an average size of ~500 bp in a Bioruptor Sonicator (Diagenode) (settings: HIGH; 30 s on; 60 s off; 10–12 cycles). Sonicated lysate was transferred to 2 ml tubes and three lysate volumes of immunoprecipitation buffer (50 mM Tris-HCl (pH 8.0); 150 mM NaCl; 5 mM MgCl; 0.5 mM EDTA; 0.2% NP-40; 5% glycerol; 0.5 mM dithiothreitol); Complete Protease Inhibitor Cocktail was added. Debris was pelleted by centrifugation for 15 min at 12,000g at 4 °C and supernatant was transferred to a new 2 ml vial. Supernatant was pre-cleared with 50 μl of Sheep-anti-Rabbit (M-280) Dynabeads (Invitrogen) for 4 h at 4 °C. Dynabeads (Invitrogen) were blocked with BSA according to the Dynabeads manual. Pre-cleared lysate was incubated with 10 μl of dynabeads suspension pre-bound for 4 h with an excess of anti-KAP1 antibody (ab10484), or anti-KRAB ZNF-antibody (ab104878). Immunoprecipitation was performed overnight at 4 °C on a rotator. Immunocomplexes were washed six times in freshly prepared RIPA buffer (50 mM Hepes (pH 8.0); 1 mM EDTA (pH 8.0); 1% (v/v) NP-40; 0.7% (w/v) deoxycholate; 0.5 M LiCl; Complete Proteinase Inhibitor Cocktail) and once in TE buffer (10 mM Tris-HCl (pH 8.0); 1 mM EDTA (pH 8.0)). H3K4me3-ChIP (H3K4me3 antibody: Milipore; catalogue no. 07-473; lot no. JBC1888194) was performed following the Roadmap Epigenome Project Protocol (April 19, 2010 version) available at http://www.roadmapepigenomics.org/protocols/type/experimental/. Immunocomplexes were eluted from the beads by incubation at 67 °C for 20 min in ChIP elution buffer (TE + 1% SDS) and vortexing every 2 min; cross-linking was reversed by incubation at 67 °C overnight. ChIP DNA was treated with RNase A/T for 2 h at 37 °C and Proteinase K for 2 h at 55 °C. NaCl was added to a final concentration of 200 mM and ChIP DNA was extracted twice with phenol/chloroform/iso-amyl-alcohol (25:24:1) and twice with chloroform/iso-amyl-alcohol (24:1). ChIP DNA was ethanol precipitated and dissolved in nuclease-free water. ChIP DNA was cleaned up one extra time using Zymo PCR purification columns.

To determine the genome-wide binding of ZNF93, we performed chromatin immunoprecipitation (ChIP) analysis, using a KRAB ZNF antibody (ab104878) which was originally raised against a peptide in ZNF486 that displays 88% identity to ZNF93 and we show is capable of recognizing ZNF93 (Extended Data Fig. 7b, c). Notably, the size of the protein immunoprecipitated by ChIP from hESC lysates corresponds to the size of ZNF93 and not ZNF486, suggesting that this antibody predominantly immunoprecipitates the highly expressed ZNF93. To establish that ZNF93 can direct ab104878 to the L1PA4 5′ UTR, ChIP-quantitative-PCR was performed on ab104878-ChIP-DNA derived from three biological replicates of TC11-mESCs transfected with either *pCAG-EV*, where EV represents an empty vector, or *pCAG-ZNF93*. Quantitative PCR was performed on a Roche LightCycler 480 II, using primers to amplify an amplicon in the 5′ UTR of L1PA4 (forward: CATTTGCGGTTCACCAATATC; reverse: GCTAGAGGTCCACTCCAGAC) and LTR12C (forward: GCACTTGAGGAGCCCTTCAG; reverse: ACACCTCCCTG CAAGCTGAG).

For ChIP-seq analysis, ChIP-DNA was used for library preparation following the Low Throughput Guidelines of the TruSeq DNA Sample Preparation kit (Illumina), with the following minor additions. Size selections were performed before and after amplification on an E-gel Safe Imager (Invitrogen) using 2% E-gel SizeSelect gels (Invitrogen). The ChIP-DNA fraction of 300–400 bp in size (including adapters) was isolated and purified. For adaptor ligations, 1 μl instead of 2.5 μl of DNA Adaptor Index was used. Indexed libraries were pooled and sequenced on the Illumina HiSEQ platform. For ChIP-seq analysis in hESCs, three biological replicates of KAP-ChIP, two biological replicates of H3K4me3-ChIP and two biological replicates of ab104878-ChIP were analysed, and for H3K4me3 ChIP-seq analysis in TC11-mESCs, two biological replicate samples were analysed for empty-vector-transfected cells and ZNF91-transfected cells, and one sample was analysed for other KZNF genes reported in Extended Data Fig. 5c. Data can be viewed on the UCSC browser: http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=http://hgwdev.soe.uc.

MACS ChiP-seq peak calling. All samples were mapped using Bowtie[28] using input Illumina fastq files consisting of paired-end reads. The human samples were mapped to the GRCh37/UCSC hg19 genome assembly. Only fully paired-end, uniquely mapping reads were kept. Potential PCR duplicates (mappings of more than one fragment with identical positions for both read ends) were removed with the samtools 'rmdup'[30] function, keeping only one of any potential duplicates. Based on the paired-end mappings, the median length of the fragments was determined for each sample. For input to MACS 1.4 (ref. 33) only the read1 mappings were used and the median

fragment length was used to determine the '-shiftsize' parameter. For each ChIP sample mappings, the corresponding input DNA sample mappings were used as a control. The UCSC table browser[34] was used to select MACS peaks that were called in both biological replicates. The overlap between KAP1 ChIP-seq replicates is ~30%, which is lower than expected and can probably be best explained by numerous retrotransposon and promoter regions on the genome displaying a low level of (possibly transient) KAP1 binding that may be below threshold in one replicate, and above threshold in the other.

Quantification of ChIP-seq and RNA-seq data for Figs 1b and 2b. For specific retrotransposon classes, the percentage of elements on human chromosome 11 (a total of 173 SVA elements; 15 full-length L1Hs elements; 84 full-length L1PA4 elements) that overlapped with KAP1 ChIP-seq peaks and H3K4me3 ChIP-seq peaks in hESCs and TC11-mESCs was determined using the UCSC table browser. Only L1PAs >5700 bp were considered to select (near) full-length L1 elements for the analysis. Transcription derived from individual SVA, full-length L1Hs and full-length L1PA4 human chromosome 11 elements in hESCs and TC11-mESCs was scored manually based on the RNA-seq coverage track uploaded in the UCSC browser, using a fixed scale that was normalized for relative sequencing depth. Level of transcription was divided in four categories: no (~0–10 reads); low (~10–30 reads); moderate (~30–50 reads) and high transcription (>50 reads). Isolated reads were not counted as transcription, nor were elements scored as transcribed when the transcription covering the retrotransposon was clearly part of exonic or intronic expression of genes. For Fig. 2b, only H3K4me3 ChIP-seq peaks that had a minimal 'score' of 100 for both empty-vector-transfected and ZNF91-transfected TC11-mESCs were considered. The 'score' is a value defined by MACS analysis representing the 'height' of each ChIP-seq signal, and the score of 100 is an arbitrary cut-off that we chose. This provides a quantitative measure of the percentage of SVAs on chromosome 11 that display a reduction of the H3K4me3 signal. For the pie charts in Fig. 3a, we used the UCSC table browser to determine the percentage of full-length L1PA elements on chromosome 11 that overlapped with an ab104878-ChIP-seq peak in the 5′ UTR (5′-most 1000 bp of each individual L1PA element). This analysis was based on 15 L1Hs, 54 L1PA2, 29 L1PA3-6030, 36 L1PA3-6160, 83 L1PA4, 39 L1PA5, 41 L1PA6, 50 L1PA7 and 14 L1PA8 full-length elements. The following should be noted about the discrepancy between the pie charts showing a small fraction of L1PA2 (7%) and L1PA7 (8%) that overlap with ab104878-ChIP-seq peaks in the 5′ UTR, and the repeat browser tracks on the left where no ab104878-ChIP-summit is observed for these elements. The annotation of L1PAs on the RepeatMasker track is based on ~500 bp in the 3′ UTR only, whereas the L1PA reference sequences in the repeat browser we used to generate the ChIP-seq summit tracks in Fig. 3a are based on the consensus of full-length L1PA sequences. In the RepeatMasker track that was used to make the pie-charts, we noticed incidental mis-annotations for these highly similar L1PA subfamilies. In particular, some L1PAs appear to be one subtype on the 3′ end (based on which they were categorized) yet are annotated as a different subfamily on the 5′ end. In fact, manual analysis of the 7% of repeat-masker-annotated L1PA2 fragments positive for KZNF-ChIP, revealed that all are mis-annotations and based on the consensus of the full length L1PA sequence should have been categorized as L1PA4 or L1PA3.

Immunoblotting. Human ESC (H9) and ZNF-transfected TC11-mESCs and HEK cells were lysed in 50 mM Tris-HCl (pH 8.0); 150 mM NaCl; 5 mM MgCl; 0.5 mM EDTA; 0.2% NP-40; 5% glycerol; 0.5 mM dithiothreitol and complete protease inhibitor cocktail (Roche) and centrifuged at max speed for 10 min at 4 °C to remove debris. Cleared lysates were subjected to SDS–PAGE on Nupage (Invitrogen) 4–12% protein gels for SDS–PAGE and transferred to nitrocellulose as described in the Nupage manual. Blots were incubated overnight in 5% non-fat dried milk in PBS-T and incubated with 1:1000 anti-KAP1 antibody (ab10484), 1:1000 anti-KZNF antibody (ab104878) or 1:1000 anti-haemagglutinin (HA; ab9110) antibody in PBS for 3 h and goat-anti-rabbit-HRP secondary antibody for 30 min at room temperature. Blots were incubated with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) and visualized on a Biorad Chemidoc MP system.

Plasmids. KZNF cDNAs were amplified from hESC cDNA, isolated from IMAGE clones or synthesized (Genscript) and cloned into pCAG EN (Addgene 11160) for transient transfections. For generation of the luciferase constructs, SVA_D (Hg19: chr11: 65, 529, 663-65, 531, 199) was synthesized (Genscript); the OCT4-enhancer region (OCT4Enh; Hg19: chr6: 31, 139, 549-31, 141, 393) was amplified by PCR from hESC gDNA, and L1PA4-5′ UTR (chr11: 74, 005, 653-74, 006, 113) was synthesized (IDT, gBlock) and were cloned upstream of a pGL4CP–SV40[34] luciferase-reporter construct. Retrotransposition assay constructs were modified from pCPE4–L1$_{RP}$–GFP[22]. Detailed plasmid descriptions and sequences of inserts can be found in Supplementary Information File 1.

Luciferase assay. Luciferase assay was carried out according to Promega dual-luciferase kit instructions and as previously published[34]. 46C[35] mESCs were plated in the afternoon on gelatin-coated 24-well plates at 35,000 cells per cm². The next morning, media was changed and 200 ng of pCAG–ZNF was co-transfected with 20 ng of SV40–luciferase reporter and 2 ng of pRL–TK–renilla (a 10:1 firefly to renilla ratio) per 24 wells using Lipofectamine2000 in duplicate wells. Twenty-four hours after transfection, wells were washed once with PBS, harvested with 100 μl of Passive Lysis Buffer for 15 min on a room-temperature rocker. Each well is then read in duplicate as 40 μl of lysate was transferred twice to a 96-well white opti-plate and combined with 50 μl of LARII substrate and read on a Perkin-Elmer luminometer and Wallace Victor Light software counting 1 s per well. Next, lysate and substrate was combined with 50 μl of Stop & Glo reagent to quench and measure renilla activity to control for transfection efficiency. Data were normalized in Microsoft Excel by dividing firefly by renilla and the average of four technical replicate measurements was taken as a raw value of activity. This activity was further normalized against an SV40–luciferase control for each KZNF pCAG construct. Final values are displayed, where for each biological replicate pCAG empty vector is set to 100%. Statistical testing was performed with a two-tailed Student's t-test and statistical differences of $P < 0.01$ are indicated in the figures. The following number of biological replicates were used: Fig. 2a: empty vector, $n = 42$; ZNF90, $n = 6$; ZNF91, $n = 17$; ZNF93, $n = 9$; ZNF254, $n = 10$; ZNF443/ZNF460/ZNF486/ ZNF519/ZNF 544/ZNF 587/ZNF589/ZNF714/ZNF721/ZNF33a, $n = 3$. Fig. 2e: empty vector, $n = 6$; human ZNF91, $n = 3$; hominine ZNF91, $n = 3$; great ape ZNF91, $n = 3$; macaque ZNF91, $n = 3$. Fig. 3c: empty vector , $n = 6$; ZNF93, $n = 3$. Fig. 3e: empty vector, $n = 6$; ZNF93, $n = 4$; ZNF93serF, $n = 6$. Extended Data Fig. 4a, $n = 6$. Extended Data Fig. 4b: no VNTR, $n = 9$; partial VNTR, $n = 3$; no hex/ Alu, $n = 2$; no hex, $n = 2$; full-length SVA, $n = 15$; SINE-R, $n = 3$. Extended Data Fig. 4c, $n = 3$. Extended Data Fig. 6c: empty vector , $n = 42$; ZNF91 (1– 11), $n = 4$; ZNF91 (1–24), $n = 7$; ZNF91 (1–30), $n = 4$; ZNF91 (1, 2, 23–36), $n = 3$. Extended Data Fig. 7a, $n = 3$. Extended Data Fig. 8b, $n = 4$.

Retrotransposition assay. The full length L1Hs retrotransposition reporter construct[22], was modified to have the 129-bp element of L1PA4 (L1Hs+129^{L1PA4}) or a scrambled 129-bp sequence (L1Hs+129scramble ^{L1PA4}) inserted at the corresponding position where the 129-bp element is present in L1PA4 and lost in L1PA3-6030. See Supplementary Information File 1 for more details on the cloning of these constructs. Retrotranspositon assay of L1Hs and related 129^{L1PA4}-containing constructs was carried out based on established protocols[22,36]. HEK293FT cells were plated at 35,000 cells per cm² on 6-well plates and incubated overnight in DMEM+FBS (without penicillin or streptomycin). The next day, cells were transfected with 300 ng of L1Hs reporter and 1 μg of pCAG–empty-vector or pCAG–ZNF93 using lipofectamine 2000/Optimem (Invitrogen); media was changed after 6 h per manufacturer recommendations. Cells were maintained and on day 4 cells were harvested with TrypLE, washed twice with PBS, placed on ice and incubated with propidium iodide. For each transfection 250,000 cells were analysed for GFP-positive and dead cells on a BD LSR II. Data were gated and analysed in FlowJo software to determine the number of live, GFP-positive cells. Statistical testing was carried out using a two-tailed Student's t-test; $n = 7$ biological replicates.

Repeat Browser. We constructed a consensus sequence of SVA_D and L1PA elements. To remove extremely short and long copies, we first eliminated the longest 2% of the copies in the genome, then took the 50 longest sequences annotated by RepeatMasker (http://www.repeatmasker.org) in the UCSC genome[37], aligned them with MUSCLE and constructed a consensus sequence from the multiple alignment. We created a version of the UCSC genome browser using this consensus as a reference sequence. MACS summits of KZNF(ab104878)-ChIP-seq and KAP1-ChIP-seq were mapped to the repeat browser for Fig. 3a, b (repeat browser: http://genome.ucsc. edu/cgi-bin/hgTracks?db=hub_27057_repeats2&position=L1PA3long%3A1-6157& hgsid=389007373_caeGCkR66TMstaDYHuKAyt6txDQD).

Multi-species *ZNF91* and *ZNF93* nucleotide sequence identification. We focused on finding homologues in other species for the fourth exon of human *ZNF91* and *ZNF93*, which contains all the important functional domains of the genes, including the KRAB domains and all the zinc-finger domains. Using BLAT from the UCSC genome browser toolset to align the human *ZNF91* (ENST00000300619) genomic nucleotide sequence (UCSC Hg19 chr19: 23, 539, 498-23, 579, 269, from 1 kb upstream to 1 kb downstream), we identified the best reciprocal hit *ZNF91* sequences in the chimpanzee (panTro4), gorilla (gorGor3), orangutan (ponAbe2), gibbon (nomLeu3), rhesus macaque (rheMac2) and baboon (papAnu2) genomes. Of note, for rhesus macaques, we used the rheMac2 assembly because we have identified a potential assembly error in the *ZNF91* fourth-exon region of the latest assembly, rheMac3, which resulted in an early stop codon. The *ZNF91* sequence obtained from rheMac2 was validated by RNA-seq data.

For *ZNF93*, the human fourth exon is located at: UCSC Hg19, chr19: 20, 043, 993-20, 045, 627. We extracted the homologous regions in other species using the UCSC 100 vertebrate species multiple sequence alignment (UCSC browser (http:// genome.ucsc.edu/), Multiz Alignments of 100 Vertebrates track). To refine the alignments, we independently aligned the human *ZNF93* fourth-exon nucleotide sequence to these homologous regions together with their immediate upstream and downstream regions (using BLAT) and manually analysed and ensured the

quality of the alignments. We obtained homologues for chimpanzee (panTro4 chr19: 20, 255, 111-20, 256, 670), gorilla (partial homologue due to missing information, gorGor3 chr19: 20, 328, 848-20, 330, 482), orangutan (partial homologue due to missing information, ponAbe2 chr19_random: 3, 818, 660-3, 820, 506), green monkey (chlSab1 chr6: 18, 428, 342-18, 430, 231), rhesus macaque (rheMac3 chr3: 73, 136, 331-73, 137, 882), crab-eating macaque (macFas5 chr19: 20, 589, 892-20, 591, 781) and baboon (papHam1 scaffold15384: 40, 473-42, 362). We aligned these sequences back to the human genome and validated that *ZNF93* was their best match. We used RAxML to construct a phylogenetic tree for these sequences and sequences of human *ZNF93* and its close relatives *ZNF90*, *ZNF737* and *ZNF626*. The results confirmed that these sequences were closest to human *ZNF93*. To check for reciprocal best matches, we aligned the human *ZNF93* fourth-exon sequence to the species genome assemblies. Due to high repetitiveness of the zinc-finger domains and high diversity of the sequences across species, the alignments resulted in a large number of matches, many of which spanned large regions (that is, false positive matches with large 'introns'). We manually analysed these alignments and confirmed that the regions listed above were the best matches.

The *ZNF93* match in gibbon (nomLeu3 chr10: 54, 583, 066-54, 586, 723) contains long insertions, indicating that there are potential errors in the gibbon reference assembly (and/or that the exon is broken into multiple exons in gibbons, and/or that the gibbon exon contains extra bases). In the next section, we explain how we used PCR to correct assembly errors in the gibbon reference to obtain a valid gibbon homologue.

Genome assembly correction at primate *ZNF91* and *ZNF93* loci. Alignments of both translated amino acid and nucleotide sequences revealed that the identified orangutan and gorilla sequences had scaffold gaps within the fourth exon of the gene *ZNF91*, which includes crucial zinc-fingers. To fill in the gaps and correct assemblies we used genomic DNA from orangutan and gorilla fibroblasts (Coriell, San Diego Zoo), and performed PCR using a selection of primers that are provided in Supplementary Information File 2. Cloned PCR products were Sanger sequenced and sequences were aligned to the corresponding assemblies as well as to the human genome using BLAT. Only clones that mapped uniquely with at least 90% coverage to the corresponding regions were kept. Similarly, orangutan and gorilla sequences had scaffold gaps within the fourth exon of the gene *ZNF93*. We used genomic DNA from Sumatran orangutan and gorilla fibroblasts (San Diego Zoo) to fill in these gaps.

We identified potential assembly errors in the gibbon reference assembly (nomLeu3). To obtain a confident homologue of the fourth exon of *ZNF93* in gibbons, we used gDNA of gibbon species *Hylobates pileatus*, *Hyloblates gabriellae* and *Nomascus leucogenys*, which were a gift from L. Carbone (Oregon Health Sciences University Primate Center) and purchased from Coriell Cell Repositories. Purified PCR products were ligated into PCR4-TOPO (Invitrogen) and sequenced. The resulting sequences were aligned to the gibbon reference assembly (nomLeu3) and were manually analysed and assembled into the consensus gibbon *ZNF93* fourth-exon sequence. The reference gibbon assembly nomLeu3 contains one tandem duplication (of the corresponding human domains 6–12) and one long insertion (~1 kb), both were refuted by sequence evidence obtained from this experiment.

Reconstructing the evolutionary history of ZNF91. Multiple sequence alignments revealed a 588-bp subsequence containing seven extra zinc-fingers in the human, chimpanzee and gorilla genomes that are not present in the orangutan, gibbon, rhesus macaque and baboon genomes. This additional sequence corresponds to zinc-fingers 6–12 of the human protein. Using BLAT to align the human copy of this sequence to the human genome, human zinc-fingers 7–12 (2–7 of the subsequence) have the best reciprocal homology to zinc-fingers 18–23 of human *ZNF91*, indicating that the subsequence was initially created by a local segmental duplication. Further analysis revealed human zinc-finger 6 (the first zinc-finger of the additional subsequence) is a near exact, best-reciprocal match of human zinc-finger 7 (the second zinc-finger of the additional sequence), indicating that after the initial intra-gene segmental duplication there was a secondary tandem duplication of the first zinc-finger. BLAT analysis revealed the additional subsequence is not present anywhere in the orangutan and other outgroup genomes. To reconstruct a parsimonious nucleotide level evolutionary history of *ZNF91*, we constructed a global multiple sequence alignment using PRANK[38] (http://www.ebi.ac.uk/goldman-srv/prank/), which simultaneously aligns the sequences and infers the ancestral sequences using a realistic model of insertion, deletion and substitution evolution. To include the two inferred duplication events in this history we created edited versions of the human, chimpanzee and gorilla sequences with the additional duplicated sequence removed and included, for each species, as two extra input nucleotide sequences, one of the first additional zinc-finger (zinc-finger 6 in the human protein), and the second of the subsequent 6 additional zinc-fingers (zinc-fingers 7–12 in the human protein). As PRANK requires a phylogenetic tree, we supplied a tree that reflects the accepted species phylogeny, but which included the additional duplications branching off after the speciation from orangutans

(Extended Data Fig. 6a). There were four amino acid changes in DNA-contacting residues in the relatively short critical time 12–8 Myr after orangutans branched off and before the human–chimpanzee–gorilla split. This together with the duplications mentioned above gives an indication of positive selection. The full multiple species alignment is shown in Supplementary Information File 3.

Reconstructing the evolutionary history of ZNF93. Multiple sequence alignment and sequence analyses (Extended Data Fig. 8a) revealed a deletion of four zinc-finger domains (located between human domains 5 and 6) in the common ancestral great ape lineage after the split with gibbons (deleted in orangutans, gorillas, chimpanzees and humans, but present in gibbons and Old-World monkeys (crab-eating macaques, rhesus macaques, baboons and green monkeys). Domains 5 and 6 (with respect to humans) are identical to each other in the great ape species. Domain 13 (with respect to humans) is missing in Old-World monkeys and is identical to domain 12 in all apes, suggesting that this domain is likely the result of a tandem duplication event that occurred in the ape last common ancestor, after the split with non-ape Old-World monkeys. Domain 17 (with respect to humans) is present in humans, crab-eating macaques and baboons (its presence or absence in rhesus macaques is unknown due to missing data), and missing in green monkeys, gibbons, orangutans, gorillas and chimps. Analysing the nucleotide sequences shows that one nucleotide insertion in the ape common ancestor (with respect to Old-World monkeys) results in an early stop codon and the loss of this domain, and a compensatory deletion of four nucleotides in humans (with respect to apes) nullifies the effect of the previous ape mutation and results in restoration of domain 17 in humans. So human ZNF93 is not like the protein of other apes. The multiple sequence alignments were obtained and validated using MUSCLE[39], MAFFT[40] and PRANK[38] and the ancestral reconstruction was constructed using PRANK. The full multiple species alignment is shown in Supplementary Information File 4.

Phylogenetic analysis and calculation of evolutionary divergence of L1PA3-6030 and L1PA3-6160 subclasses. Fifty sequences of L1PA3-6030, 50 sequences of L1PA3-6160, 3 sequences of L1PA2 and 3 sequences of L1PA4 were aligned by ClustalW in MEGA6 software package[41]. Only full-length L1PAs were selected. For phylogenetic analysis, the sequence downstream of the 129-bp element (L1PA4 and L1PA3-6160), or the corresponding position (L1PA2 and L1PA3-6030) was used to generate phylogenetic trees. Multiple methods were used (Maximum Parsimony, Minimum Likelihood and Minimum Evolution) to generate trees with comparable outcome. The phylogenetic tree generated by the Minimum Evolution method[42] was used to calculate the divergence times for all branching points with the RelTime method[43].

To calculate the average divergence from consensus, first consensus sequences were calculated for L1PA3-6030 and L1PA3-6160 from 150 full-length elements of each subclass using EMBOSS software (http://www.emboss.sourceforge.net/). Each consensus sequence was aligned in MEGA6 with the respective 150 full-length element by ClustalW. In order to be able to compare values for L1PA3-6030 and L1PA3-6160 to divergence values for other L1PA subfamilies, determined previously[9], we used the 500 bp of the 3′ end of the L1PA3 subclasses, and excluded the poly(A)-stretch at the 3′ end of L1PAs. The pairwise distances for each of the 151 (500 bp) sequences (150 individual L1PAs and 1 consensus) were calculated in MEGA6 and plotted in a distance matrix. The average distance (divergence) from consensus was determined by calculating the mean distance (± s.e.m.) from the consensus sequence to each individual L1PA3 element. The age of each L1PA3 subclass was estimated using a base-pair substitution rate of 0.17% per million years (Myr)[9].

VNTR size analysis for SVA-subfamilies. We extracted RepeatMasker SVA elements in the human genome as annotated in the UCSC Genome Browser RepeatMasker track (Hg19.rmsk). Each element was annotated with Tandem Repeat Finder[44] to identify all base pairs covered by a tandem repeat. While VNTR and HEX domains are both tandem repeats, we assumed that the length of the HEX region is a lot shorter and relatively fixed compared to the VNTR, so in the following we use the length of all base pairs masked by Tandem Repeat Finder as a proxy for the length of the VNTR. SVAs annotated by RepeatMasker as multiple adjacent SVA fragments can correspond to a single full-length SVA element. Therefore, to restrict our analysis to unbroken full-length elements, we concentrated on elements that displayed an intact SVA structure, with at least 800 bp of sequence outside of the VNTR region, a size that corresponds to the sizes of Alu and SINE-R combined. For this enriched set of SVAs the histogram of VNTR lengths is plotted in Extended Data Fig. 10c.

Determination of changes per million years for Fig. 4. For ZNF91 and ZNF93, we counted the numbers of zinc-fingers that have undergone structural changes that could affect DNA binding specificity for each of the evolutionary branchpoints, based on the multiple sequence analysis and ancestral reconstruction (see Methods sections 'Reconstructing the evolutionary history of ZNF91' and 'Reconstructing the evolutionary history of ZNF93'). Changes in DNA binding residues, zinc-finger

deletions or zinc-finger duplications/gains were all weighted equally and counted as '1' because it is unpredictable how each of these changes may change target DNA recognition. The number of changes from one branchpoint to another was divided by the number of million years of that timeframe to determine the number of zinc-fingers that changed per million years. For zinc-fingers in ZNF93 that were different between macaques and gibbons, but conserved between gibbons and great apes, we lacked an outgroup species necessary to determine when the changes occurred. Therefore, to get a rough estimate, we divided the total number of changes between macaques and gibbons, by the amount of time on each of these lineages. From the point of divergence of Old-World monkeys to present-day macaques is 25 Myr, from the point of divergence of Old-World monkeys to the LCA of gibbon and great apes is 7 Myr (25–18 Myr). Therefore we estimated that about 75% of the observed changes happened on the macaque lineage and 25% of the changes on the lineage to the LCA of gibbons and great apes. Similarly, for L1PA elements the consensus sequences of each L1PA element was compared to its direct predecessor and successor, and base-pair substitutions, deletions or insertions were all counted as '1'. The number of base-pair changes per site within the 5′ UTR (1,000 bp) from one L1PA element and its successor was divided by the number of years within the time-frame each L1PA-subfamily was dominant[9]. (See Methods section 'Phylogenetic analysis and calculation of evolutionary divergence of L1PA3-6030 and L1PA3-6160 subclasses') to get the base-pair changes per site per Myr values. For SVA, the percentage of VNTR increase per Myr between SVA-subfamilies is indicated for the timeframe from the emergence of one SVA subfamily to the successor. The average VNTR size for SVA-subtypes as determined in this study (Extended Data Fig. 10c) and the estimated time-points of emergence previously reported for SVA-subfamilies[12] were used to calculate the percentage increase of VNTR size per Myr.

26. Parkhomchuk, D. *et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37,** e123 (2009).
27. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
28. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
29. Hsu, F. *et al.* The UCSC known genes. *Bioinformatics* **22,** 1036–1046 (2006).
30. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
32. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
33. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
34. Onodera, C. S. *et al.* Gene isoform specificity through enhancer-associated antisense transcription. *PLoS ONE* **7,** e43511 (2012).
35. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnol.* **21,** 183–186 (2003).
36. Hancks, D. C., Mandal, P. K., Cheung, L. E. & Kazazian, H. H. The minimal active human SVA retrotransposon requires only the 5′-hexamer and Alu-like domains. *Mol. Cell. Biol.* **32,** 4718–4726 (2012).
37. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
38. Löytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320,** 1632–1635 (2008).
39. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
41. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).
42. Rzhetsky, A. & Nei, M. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **9,** 945–967 (1992).
43. Tamura, K. *et al.* Estimating divergence times in large molecular phylogenies. *Proc. Natl Acad. Sci. USA* **109,** 19333–19338 (2012).
44. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
45. Naas, T. P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* **17,** 590–597 (1998).
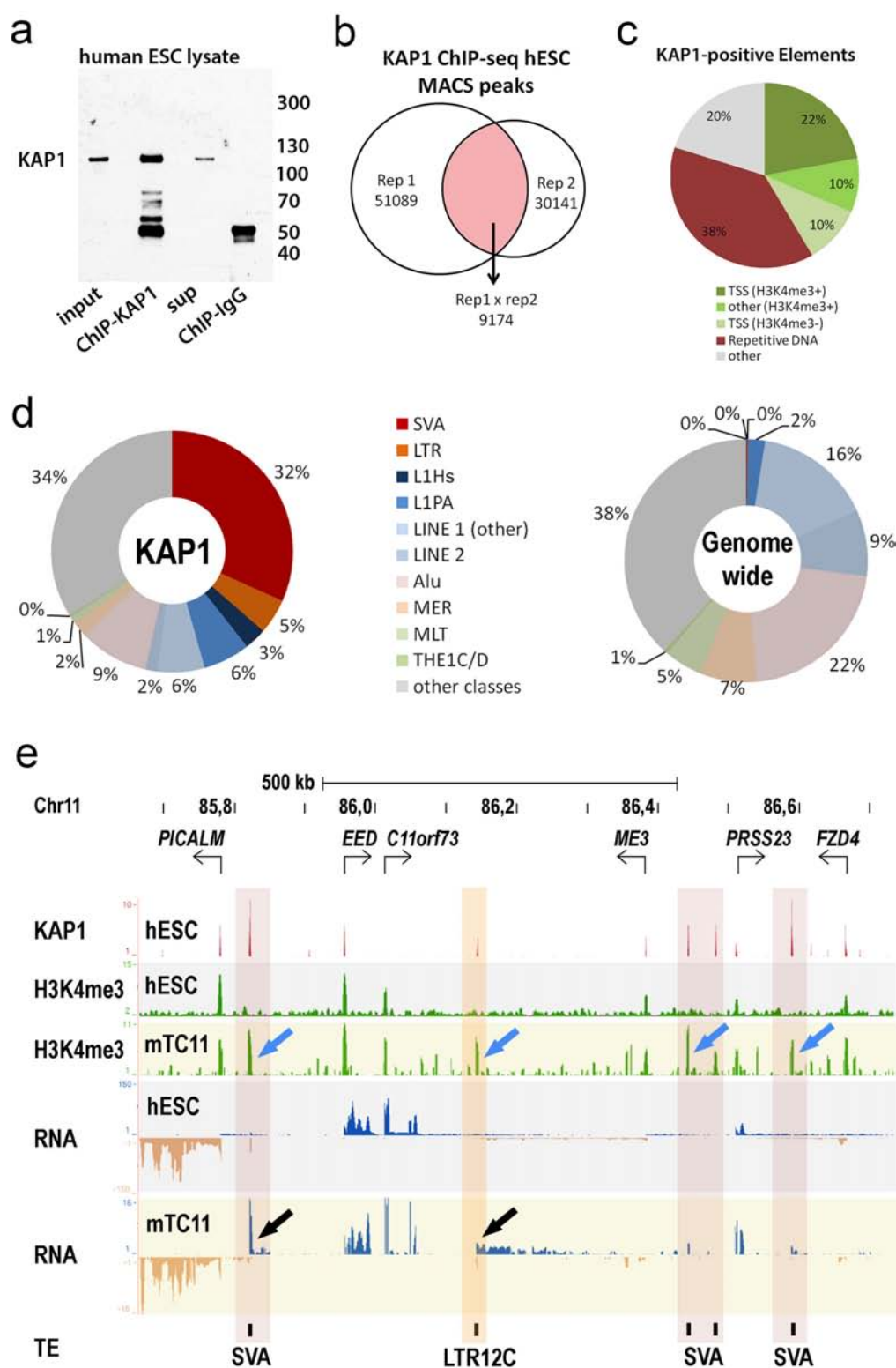
Extended Data Figure 1 | KAP1 associates with recently emerged transposable elements. a, Immunoblot incubated with anti-KAP1 antibody loaded with 1% input and eluates of KAP1-ChIP or IgG-ChIP derived from hESC lysates. **b**, Diagram showing numbers of KAP1 peaks identified in two independent biological replicates and common peaks. **c**, Distribution of 9,174 KAP1-ChIP-seq peaks over various DNA elements. **d**, Distribution of retrotransposon classes among KAP1-ChIP peaks from hESCs (left) or genome-wide (right). **e**, KAP1 and H3K4me3 ChIP-seq and RNA-seq coverage tracks for a representative region on human chromosome 11 in hESCs (white- or grey-shaded) and TC11-mESCs (yellow-shaded). Blue arrows, derepressed retrotransposons; black arrows, re-activated transcription; red vertical shading, reactivated SVAs; orange shading, reactivated LTR12C. Blue and tan RNA-seq tracks indicate positive and negative strand transcripts, respectively. Note that while the majority of SVAs display aberrant H3K4me3 signal, for unclear reasons not all SVAs display aberrant transcription in TC11-mESCs. Rep, biological replicate; sup, supernatant; TSS, transcription start site.
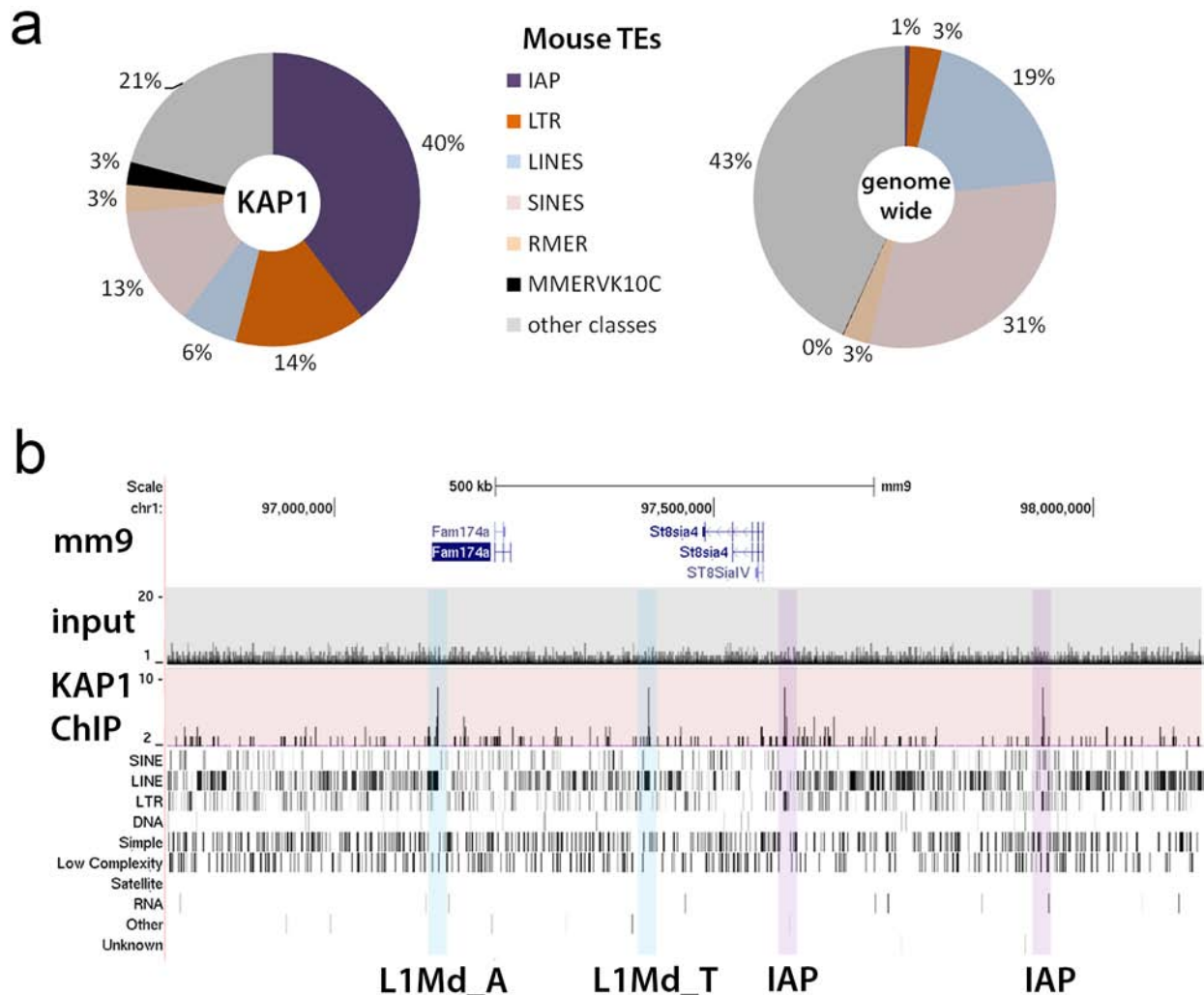
Extended Data Figure 2 | Mouse KAP1 associates with mouse-specific retrotransposons in mouse ESCs. **a**, Distribution of KAP1-ChIP-Seq reads from mESCs (left) and the mouse genome (right) for retrotransposon families as defined by RepeatMasker (http://www.repeatmasker.org/). **b**, UCSC Browser image displaying ChIP-seq tracks for input (grey shading) and KAP1 (red shading) as well as gene annotation and repeat element tracks for a region on mouse chromosome 1. Blue shading, KAP1-positive active mouse L1-subtypes[45]; purple shading, KAP1-positive active intracisternal A-particle (IAP) retrotransposons. LINES, long interspersed nuclear elements; LTR, long terminal repeat; MMERVK10C, mouse endogenous retrovirus subtype K10C; RMER, medium reiteration frequency repetitive sequence; SINES, short interspersed nuclear elements; TEs, transposable elements.
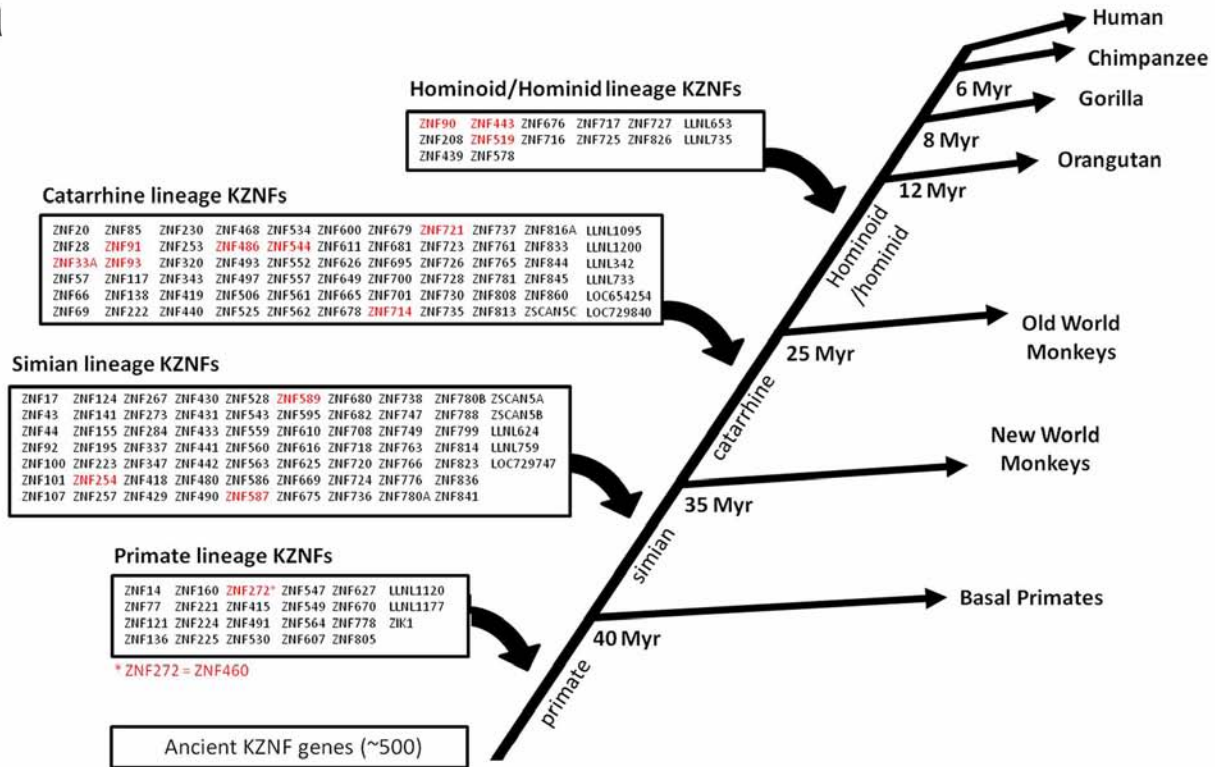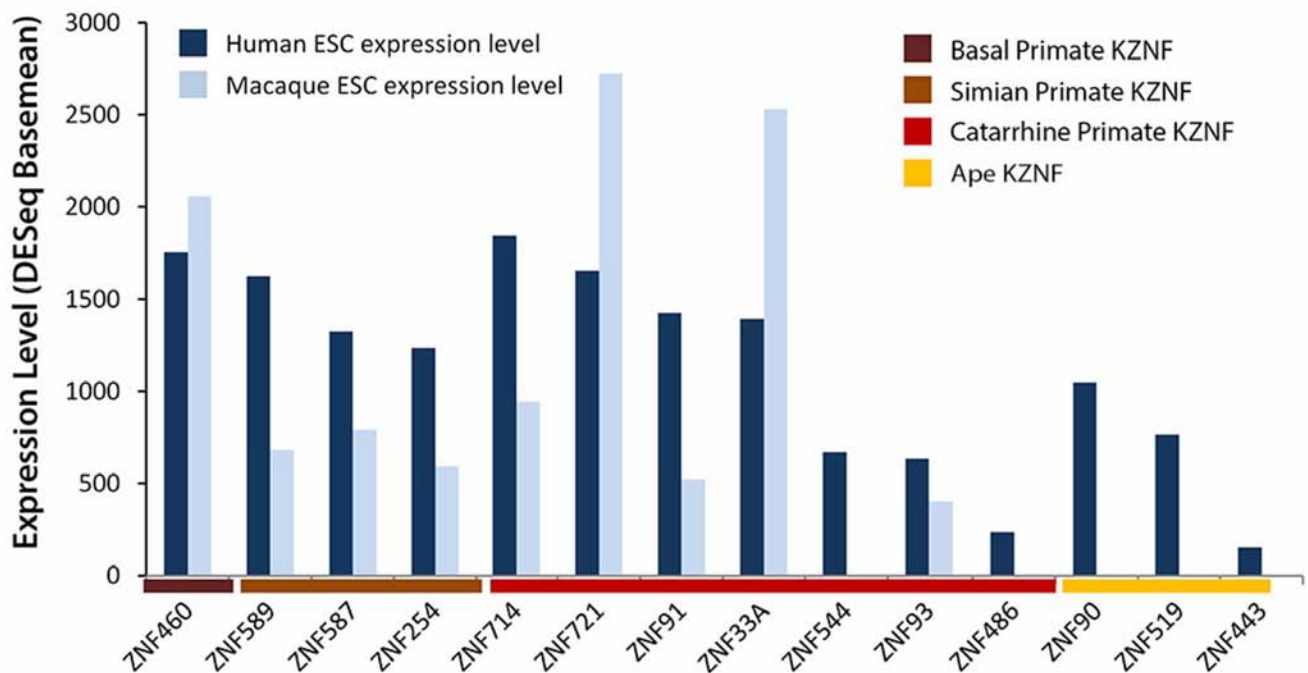
Extended Data Figure 3 | Selection of primate-specific KZNF genes with high expression in hESCs. a, Schematic of primate-specific KRAB zinc-finger genes subdivided in different clades based on previous analysis[7]. KZNFs shown in **b** are highlighted in red. **b**, DESeq-calculated gene expression levels for the 17 highest expressed KRAB zinc-finger genes in hESCs (dark blue) and macaque ESCs (light blue), subdivided by clades.
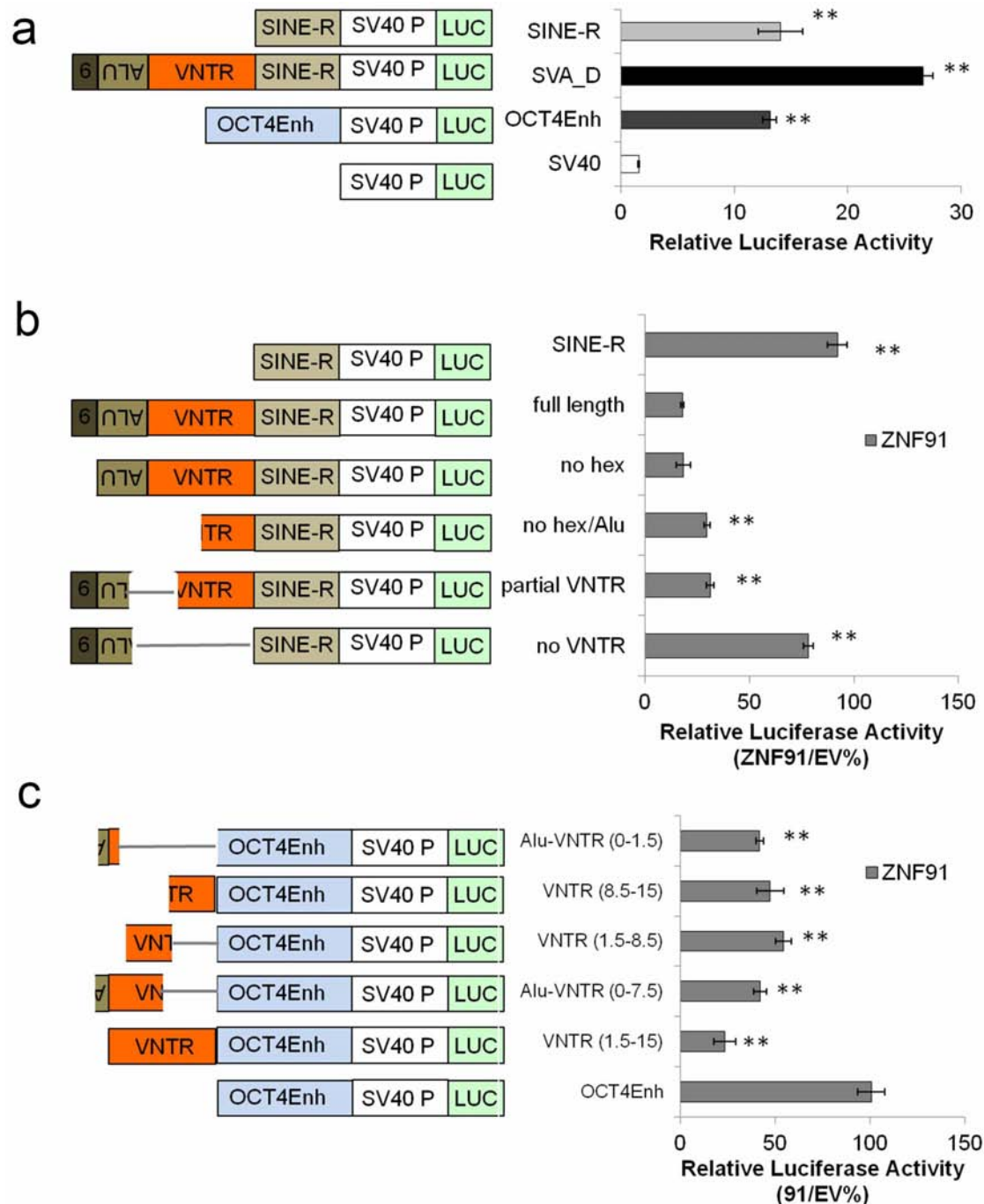
Extended Data Figure 4 | The SVA VNTR domain is necessary and sufficient for ZNF91-mediated repression of luciferase activity.
a–c, Schematic of SV40–luciferase constructs used (left) and relative luciferase activity after transfection of the indicated constructs in mESCs (right). **a**, SVA and SINE-R are strong enhancers ($n = 6$ biological replicates). **b**, Deletion analysis reveals the VNTR of SVA is required for ZNF91-mediated reporter regulation. Luciferase activity in the presence of ZNF91 expressed as a ratio of that observed for empty vector with the same reporter. Biological replicates: no VNTR, $n = 9$; partial VNTR, $n = 3$; no hex/Alu, $n = 2$; no hex, $n = 2$; full length SVA, $n = 15$; SINE-R, $n = 3$. Empty vector is set to 100% for comparison. **c**, 1.5 VNTR repeats are sufficient to confer ZNF91-mediated regulation on an OCT4Enh–SV40–luciferase-reporter. $n = 3$ biological replicates. **$P < 0.01$; error bars are s.e.m.
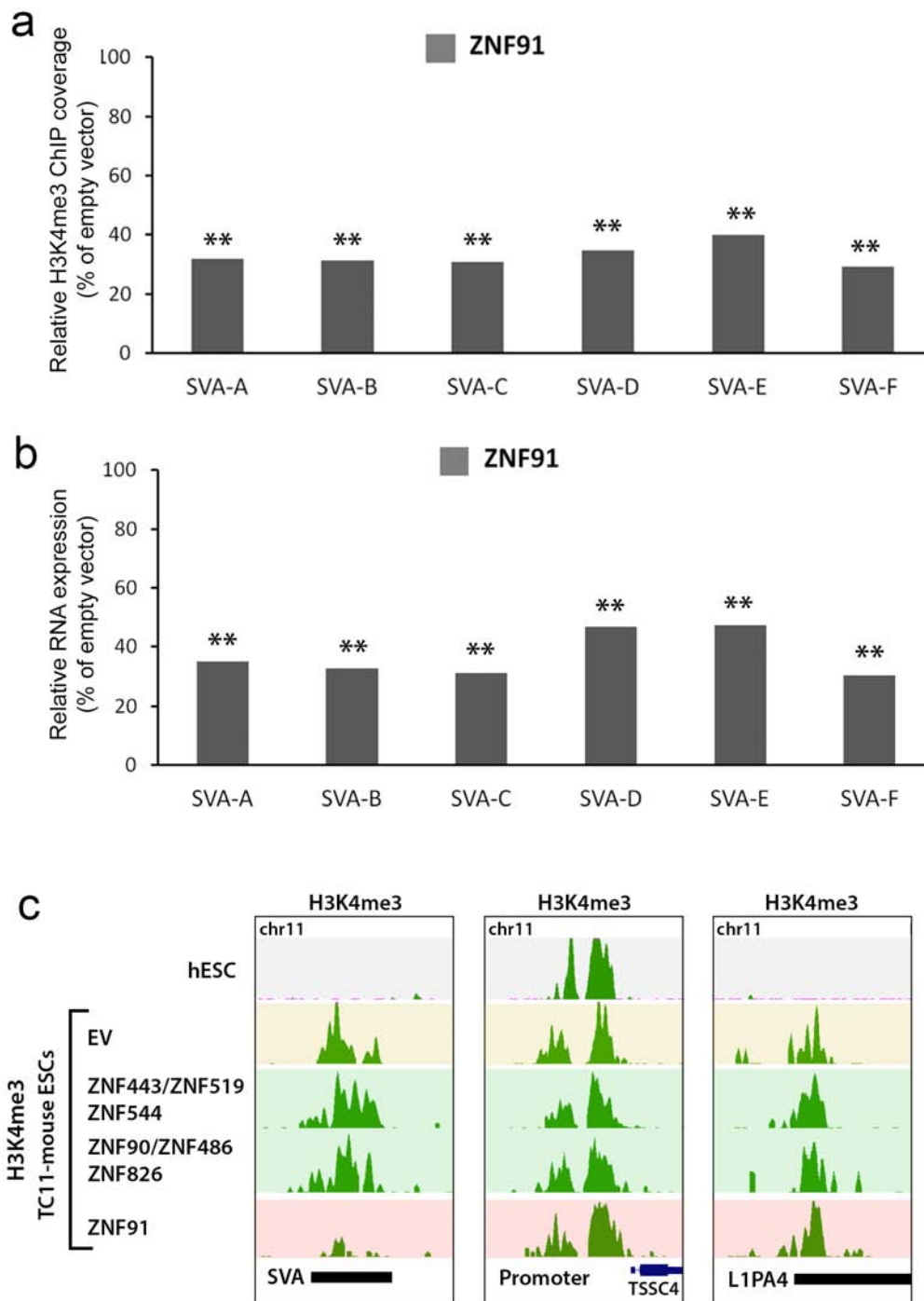
Extended Data Figure 5 | SVA is specifically repressed *in vivo* by ZNF91. a, b, Normalized DESeq basemean values for H3K4me3 ChIP-seq (**a**) and RNA-seq (**b**) for retrotransposon classes that showed a significant change in ZNF91-transfected TC11-mESCs relative to empty vector. SVAs were the only transposable elements that showed a significant decrease in H3K4me3 and RNA-seq values. *\*Benjamini–Hochberg adjusted-$P < 0.01$. **c**, UCSC browser images for a representative SVA element, promoter and L1PA4 element, showing H3K4me3 ChIP-seq signal for hESCs (grey), TC11-mESCs transfected with empty vector (yellow), pools of primate-specific KRAB zinc-fingers (green) and ZNF91 (red). TSSC4: tumor-suppressing subtransferable candidate 4.
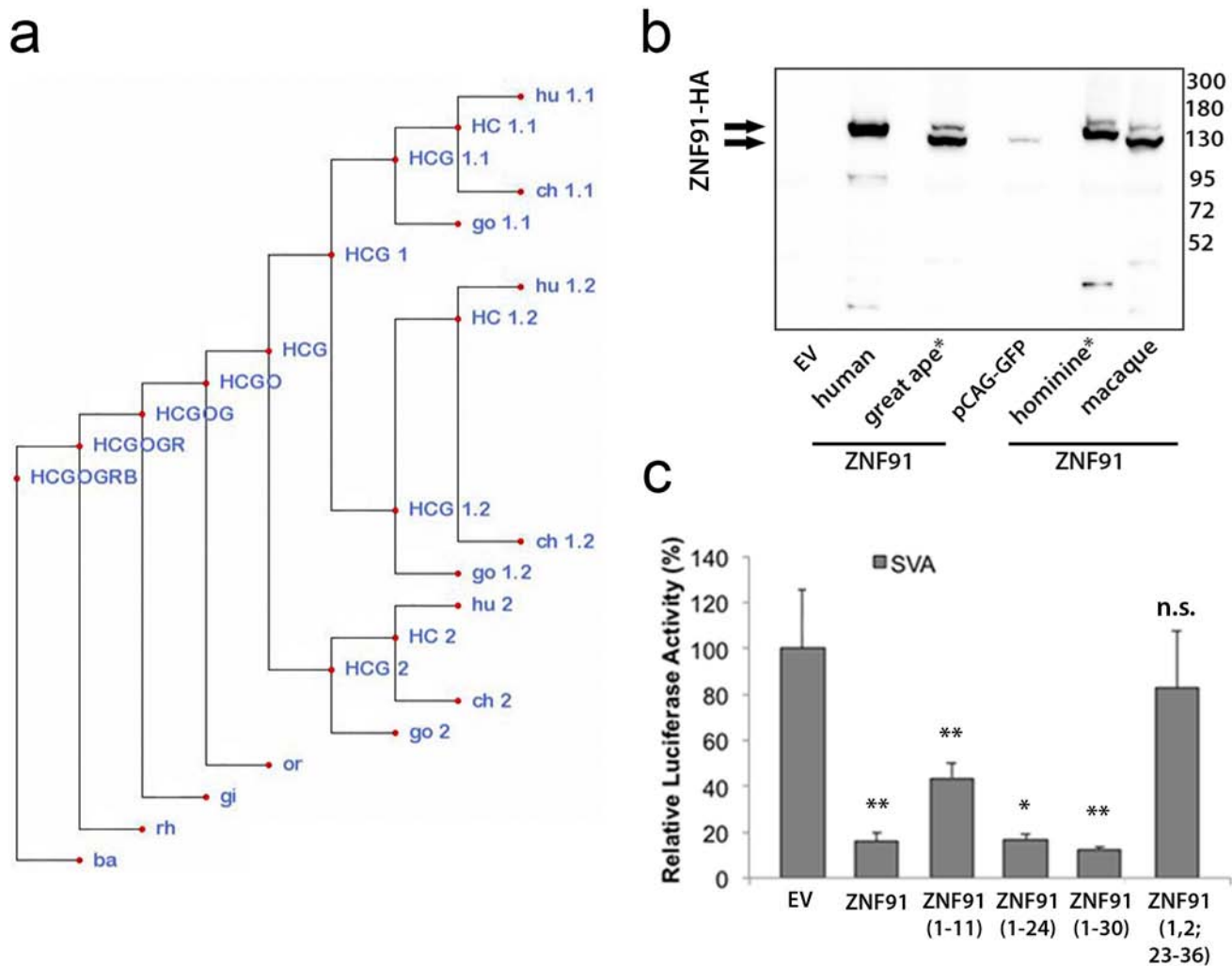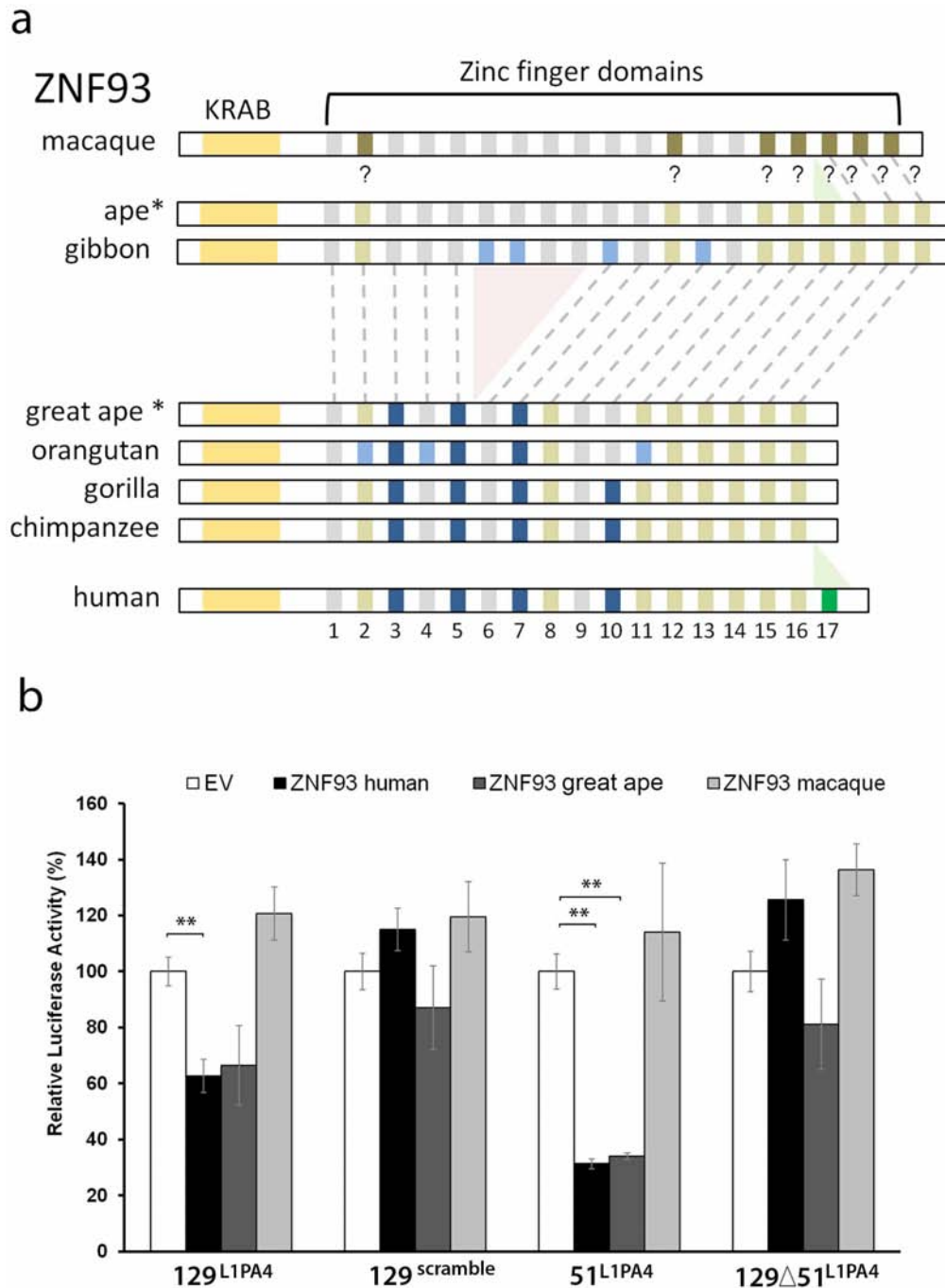
Extended Data Figure 6 | Evolutionary history of ZNF91. a, The phylogenetic tree used in multiple sequence alignment and ancestral reconstruction of ZNF91 (Supplementary Information File 3). 'hu 1.1', 'ch 1.1' and 'go 1.1' represent human, chimpanzee and gorilla domain 6, respectively, 'hu 1.2', 'ch 1.2', 'go 1.2' represent human, chimpanzee and gorilla domains 7–12, respectively, and 'hu 2', 'ch 2' and 'go 2' represent the ZNF91 sequence from start to domain 5, a breakpoint, and from domain 13 to the end (see Methods). Ancestors are labelled with first letters of leaf species below them, for example, HCG is a human–chimp–gorilla ancestor. **b,** Immunoblot incubated with anti-HA antibody on lysates of HEK293FT cells transfected with HA-tagged human, great ape, hominine and macaque ZNF91 proteins or lysates transfected with an empty vector and pCAG–GFP. Asterisks denote reconstructed ancestral proteins. **c,** ZNF91 domain deletion analysis showing relative luciferase activities on the SVA-D–SV40 luciferase reporter after transfection of empty vector or ZNF91 deletion constructs in mESCs. Error bars are standard deviation. Numbers in parenthesis indicate zinc-fingers present in the ZNF91 deletion construct. $*P < 0.05$; $**P < 0.01$. Biological replicates: empty vector, $n = 42$; ZNF91 (1–11), $n = 4$; ZNF91 (1–24), $n = 7$; ZNF91 (1–30), $n = 4$; ZNF91 (1, 2, 23–36), $n = 3$.
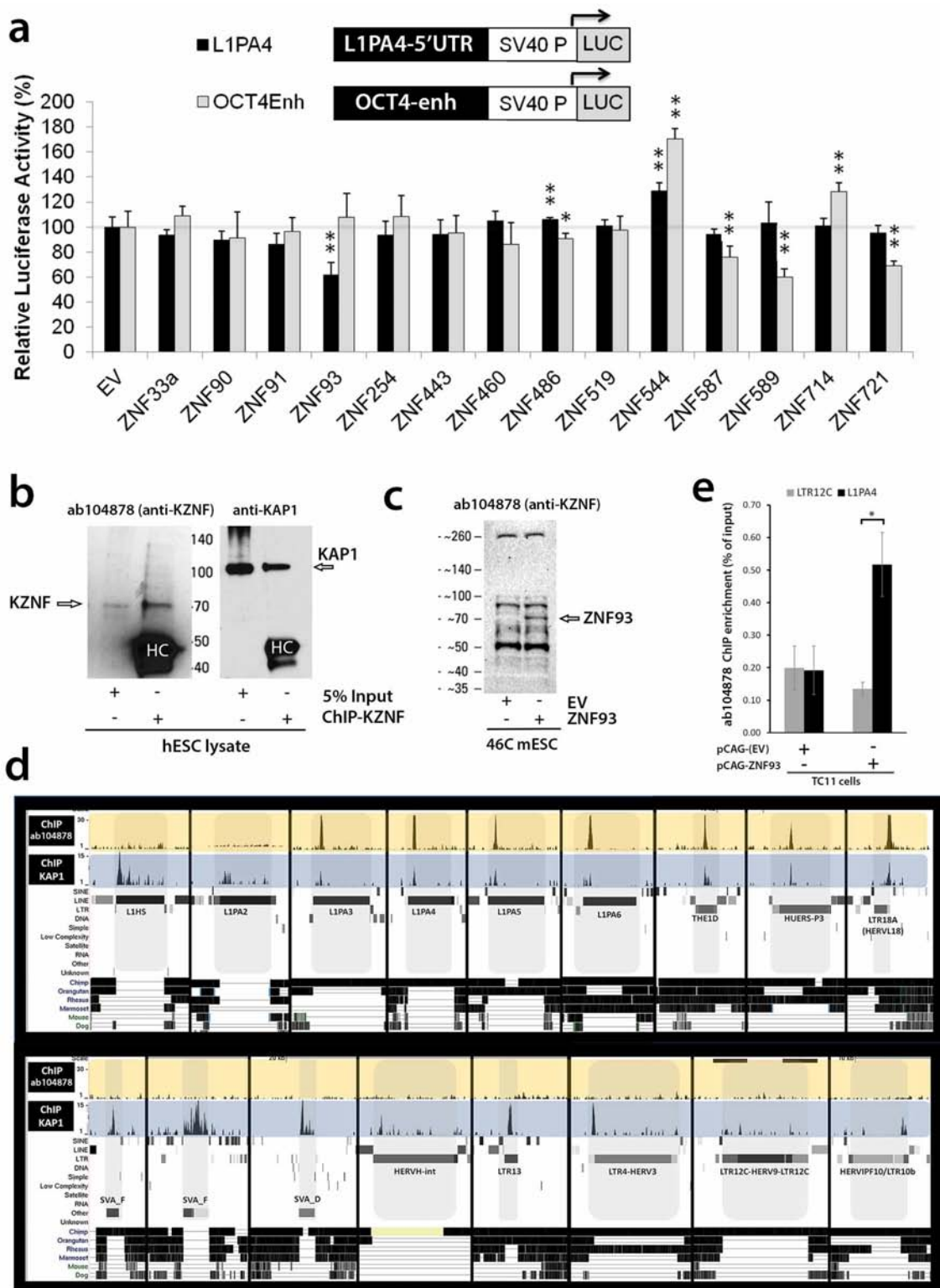
Extended Data Figure 7 | L1PA4 elements are repressed by primate-specific ZNF93. a, Relative luciferase activity on a L1PA4– and a OCT4-enhancer– SV40–luciferase-reporter after transfection of 14 KZNFs in mESCs. Significance measured relative to empty vector. $n = 3$ biological replicates; *$P < 0.05$; **$P < 0.01$; error bars are s.e.m. **b**, Immunoblot showing that ChIP with antibody ab104878 predominantly reacts with a protein of ~70 kDa (left panel) and co-immunoprecipitates KAP1 (right panel). HC, heavy chain of IgG. **c**, Immunoblot demonstrating that ChIP with ab104878 detects overexpressed ZNF93 in 46c mESCs as a ~70 kDa protein. **d**, Repeat Browser (see Methods) displaying ChIP-seq coverage tracks for ab104878 (ZNF93; yellow shading) and KAP1 (blue shading) for a selection of KAP1-bound retrotransposons. **e**, ChIP-qPCR for amplicons in L1PA4 and LTR12C elements on chromosome 11 in TC11-mESCs after transfection with an empty vector or ZNF93 and ChIP with ab104878. ChIP enrichment is plotted as percentage of input. $n = 3$ biological replicates; *$P < 0.05$; error bars are s.e.m.

Extended Data Figure 8 | Reconstruction of the evolutionary history of ZNF93. **a**, Schematic based on the multiple sequence alignment of ZNF93 orthologues (Supplementary Information File 4). Red shaded area, deletion of zinc-fingers; green shaded area, gain of zinc-fingers; green stripes, gained zinc-fingers; dark blue stripes, zinc-fingers that changed contact residues in the lineage to humans; light blue stripes, changes in other lineages; brown stripes, zinc-fingers with different binding residues between macaques and gibbons, with gibbons sharing the great ape conformation. For this last group of zinc-fingers, it is unknown (represented with a ? symbol) whether the change happened in monkeys or in the LCA of gibbons and great apes after the divergence of Old-World monkeys (see Methods). Asterisks denote reconstructed ancestral proteins. **b**, Relative OCT4-enhancer–SV40p–luciferase activity for reporters with the indicated L1PA4-derived sequences after co-transfection of an empty vector or various ZNF93 constructs. **P < 0.01; error bars are s.e.m.
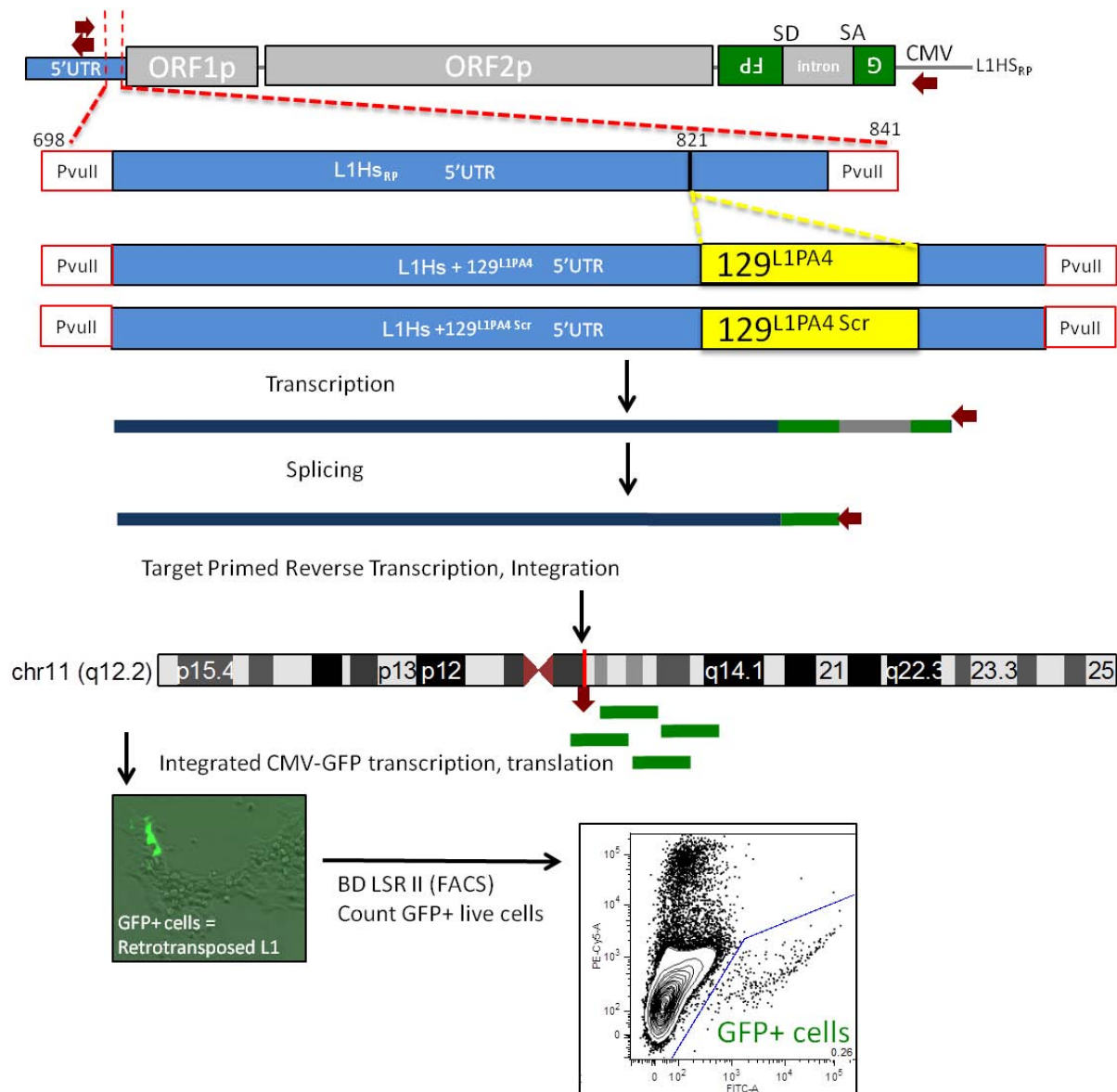
Extended Data Figure 9 | Schematic of L1Hs retrotranspostion assay.
a, Schematic of constructs tested indicating the site of 129[L1PA4] transplant into L1Hs and concept of L1–GFP assay[24] in which GFP expression marks cells where a transfected L1 episome has retrotransposed into a HEK293 cell's chromosomes. ORF, open reading frame; CMV, cytomegalovirus promoter; SD, splice donor; SA, splice acceptor; PvuII, restriction enzyme site.
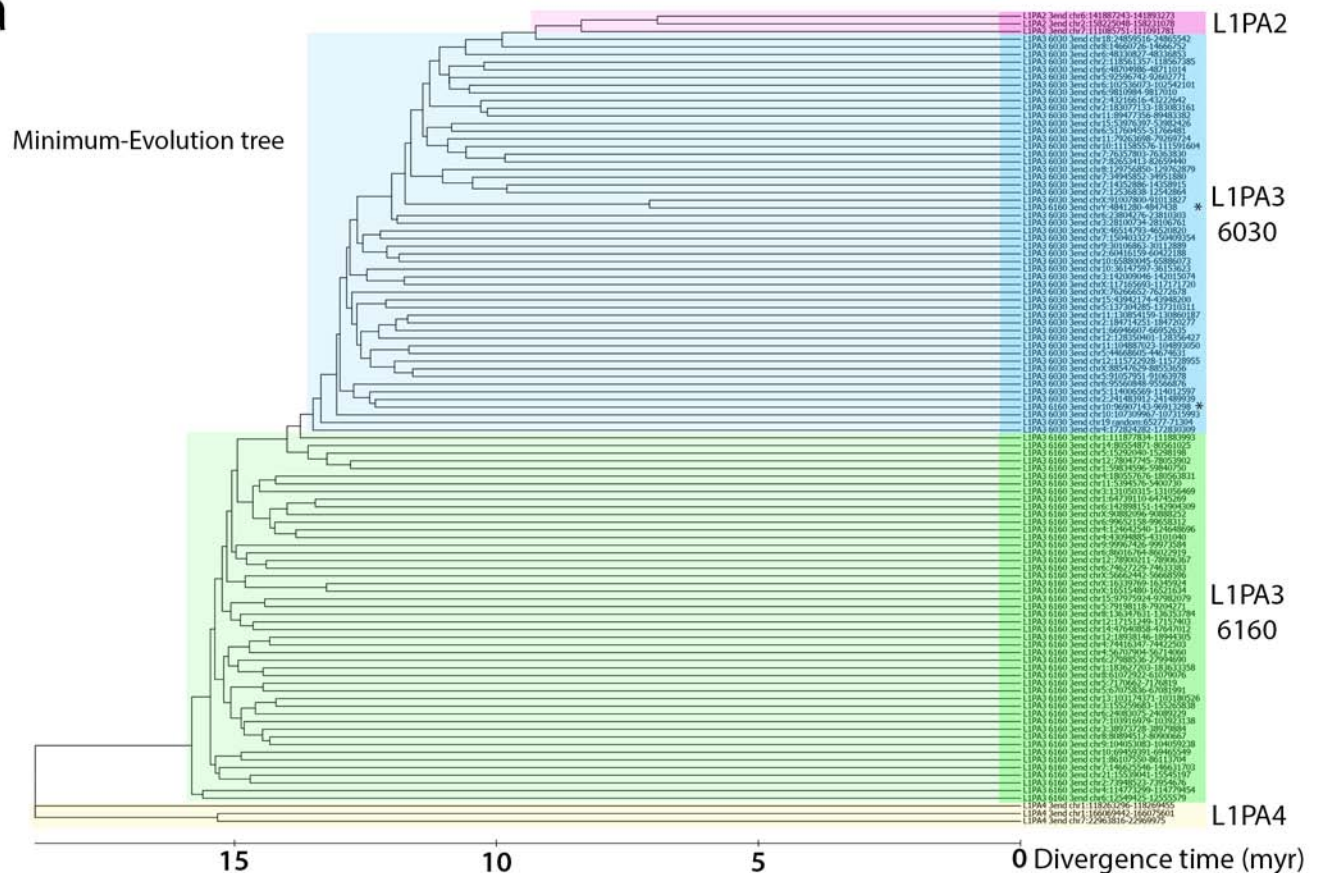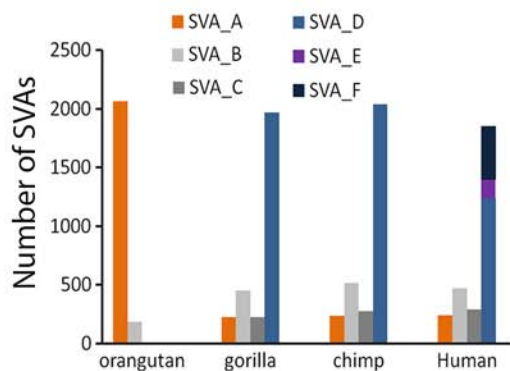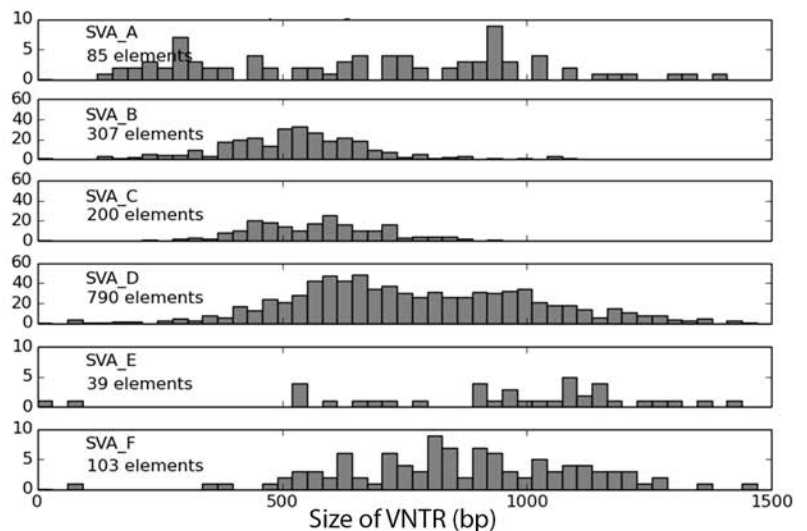
Extended Data Figure 10 | Evolutionary history of L1PA3-6030, L1PA3-6160 and the VNTR size in SVA. a, Phylogenetic tree, rooted on L1PA4, generated using the Minimum Evolution method[42] for fifty 3′-end sequences of L1PA3-6030 and L1PA3-6160, and three 3′-end sequences for L1PA2 and L1PA4. **b**, Bar graphs showing the number of SVA-_A through SVA_F insertions in each great ape genome. **c**, Distribution of VNTR size for untruncated SVA elements in the human genome plotted for each SVA-subfamily. The number of untruncated elements identified for each subtype is indicated.

LETTER

Dietary modulation of the microbiome affects autoinflammatory disease

John R. Lukens[1], Prajwal Gurung[1], Peter Vogel[2], Gordon R. Johnson[1], Robert A. Carter[3], Daniel J. McGoldrick[3], Srinivasa Rao Bandi[1], Christopher R. Calabrese[4], Lieselotte Vande Walle[5,6], Mohamed Lamkanfi[5,6] & Thirumala-Devi Kanneganti[1]

The incidences of chronic inflammatory disorders have increased considerably over the past three decades[1]. Recent shifts in dietary consumption may have contributed importantly to this surge, but how dietary consumption modulates inflammatory disease is poorly defined. *Pstpip2*[cmo] mice, which express a homozygous Leu98Pro missense mutation in the *Pombe* Cdc15 homology family protein PSTPIP2 (proline-serine-threonine phosphatase interacting protein 2), spontaneously develop osteomyelitis that resembles chronic recurrent multifocal osteomyelitis in humans[2–4]. Recent reports demonstrated a crucial role for interleukin-1β (IL-1β) in osteomyelitis, but deletion of the inflammasome components caspase-1 and NLRP3 failed to rescue *Pstpip2*[cmo] mice from inflammatory bone disease[5,6]. Thus, the upstream mechanisms controlling IL-1β production in *Pstpip2*[cmo] mice remain to be identified. In addition, the environmental factors driving IL-1β-dependent inflammatory bone erosion are unknown. Here we show that the intestinal microbiota of diseased *Pstpip2*[cmo] mice was characterized by an outgrowth of *Prevotella*. Notably, *Pstpip2*[cmo] mice that were fed a diet rich in fat and cholesterol maintained a normal body weight, but were markedly protected against inflammatory bone disease and bone erosion. Diet-induced protection against osteomyelitis was accompanied by marked reductions in intestinal *Prevotella* levels and significantly reduced pro-IL-1β expression in distant neutrophils. Furthermore, pro-IL-1β expression was also decreased in *Pstpip2*[cmo] mice treated with antibiotics, and in wild-type mice that were kept under germ-free conditions. We further demonstrate that combined deletion of caspases 1 and 8 was required for protection against IL-1β-dependent inflammatory bone disease, whereas the deletion of either caspase alone or of elastase or neutrophil proteinase 3 failed to prevent inflammatory disease. Collectively, this work reveals diet-associated changes in the intestinal microbiome as a crucial factor regulating inflammasome- and caspase-8-mediated maturation of IL-1β and osteomyelitis in *Pstpip2*[cmo] mice.

Changes in diet are known to determine susceptibility to common autoimmune diseases such as atherosclerosis, coronary heart disease and type II diabetes[7]. To address whether dietary intake affects osteomyelitis in *Pstpip2*[cmo] mice, a cohort of animals were fed *ad libidum* a diet rich in high saturated fats and cholesterol (high-fat diet, or HFD), and disease progression was compared to that of *Pstpip2*[cmo] mice placed on a regular low-fat diet (LFD). As expected, all animals on a LFD ($n = 40$) had developed inflammatory bone disease by day 100 (Fig. 1a), as evidenced by the red and swollen appearance of their hind paws (Extended Data Fig. 1a), the significant bone erosion and deformity seen in representative isosurface micro-computed tomography micrographs (Fig. 1b), and the increased size of draining popliteal lymph nodes (Extended Data Fig. 1b). In marked contrast, *Pstpip2*[cmo] mice that were fed a HFD ($n = 22$) were largely protected from osteomyelitis, and these mice resembled healthy wild-type mice in terms of hind paw inflammation, bone erosion and lymph node size (Fig. 1a, b and Extended Data Fig. 1b). In
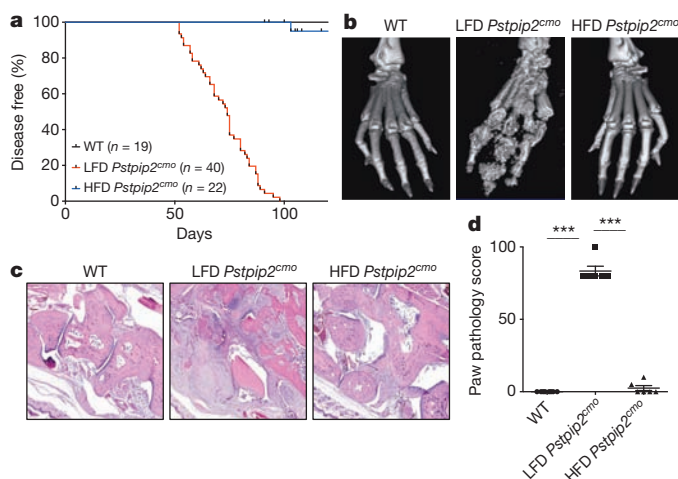
Figure 1 | Changes in diet limit the development of inflammatory bone disease in *Pstpip2*[cmo] mutant mice. a–d, Wild-type (WT) and *Pstpip2*[cmo] mutant mice were fed a low-fat diet (LFD) or a high-fat and cholesterol diet (HFD). **a**, Incidence of inflammatory bone disease. Combined data from three independent experiments. **b–d**, Representative isosurface micro-computed tomography paw scans (**b**), haematoxylin and eosin sections (original magnification, ×4) (**c**) and pathology scores (**d**) for hind paw samples from 12–14-week-old wild-type, LFD *Pstpip2*[cmo] and HFD *Pstpip2*[cmo] mice. Each point represents an individual mouse, and the line represents the mean ± s.e.m. ***$P < 0.001$; Student's *t*-test.
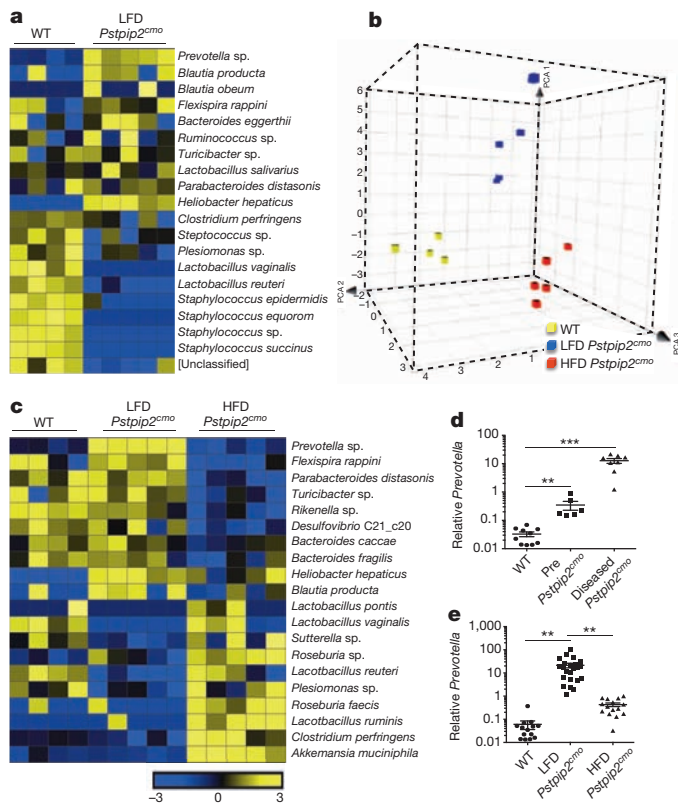
agreement, haematoxylin and eosin-stained sections of the hind paws and tails of *Pstpip2*[cmo] mice that were fed a HFD were devoid of infiltrating inflammatory cells and lacked signs of osteolytic bone destruction (Fig. 1c, d and Extended Data Fig. 1c, d). Conversely, *Pstpip2*[cmo] mice that were fed a regular LFD diet showed significant bone destruction and inflammatory cell infiltration in stained paw (Fig. 1b–d) and tail (Extended Data Fig. 1c, d) sections. In agreement, profound reductions in the numbers of infiltrating neutrophils and macrophages were evident in the footpads of HFD-fed *Pstpip2*[cmo] mice compared to LFD-fed *Pstpip2*[cmo] mice (Extended Data Fig. 1e). Consumption of a HFD was also found to rescue hyperinflammatory cytokine production in *Pstpip2*[cmo] mutant mice (Extended Data Fig. 2a, b). As expected for mice on a BALB/cJ genetic background[8], *Pstpip2*[cmo] mice retained a normal body weight during these studies, regardless of whether they were fed a lean or high-fat diet (Extended Data Fig. 3a, b). Collectively, these observations demonstrate that the dietary composition determines to a large extent whether genetically susceptible *Pstpip2*[cmo] mice develop osteomyelitis independently of gross changes in body weight.

Diets high in fat and cholesterol induce large-scale changes in the host microbiota composition[9,10]. We made use of 16S ribosomal RNA (rRNA)

[1]Department of Immunology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [2]Animal Resources Center and the Veterinary Pathology Core, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [3]Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [4]Small Animal Imaging Core, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [5]Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium. [6]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium.
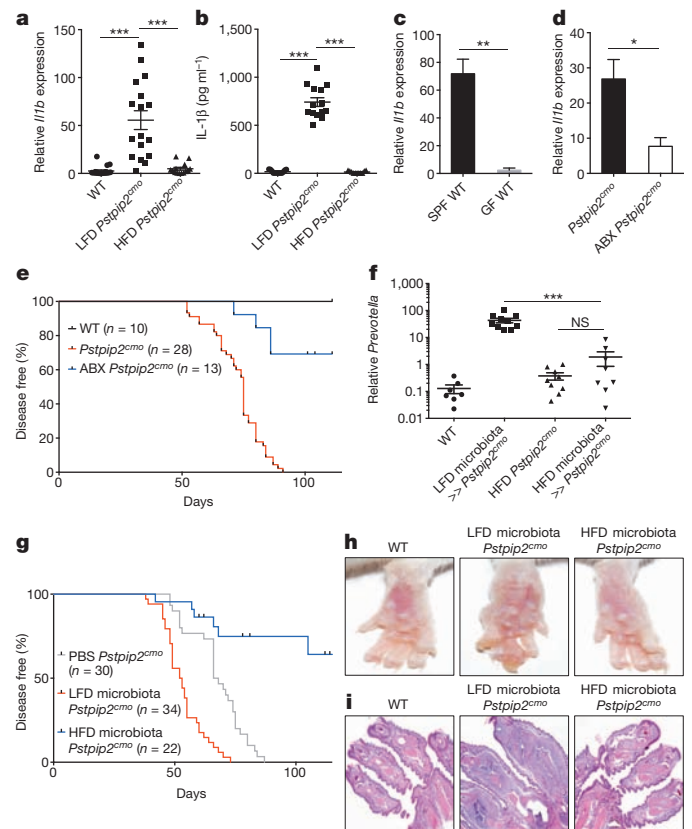
metagenomic sequencing to address whether inflammatory bone disease in *Pstpip2^cmo* mice was associated with intestinal dysbiosis that was rescued by a HFD regimen. The commensal intestinal ecology of *Pstpip2^cmo* mice that were fed a regular LFD was markedly different from the microbiota of healthy age- and sex-matched wild-type mice (Fig. 2a). Notable alterations included the outgrowth of *Prevotella* and concomitant reductions in *Lactobacillus* genera in LFD-fed *Pstpip2^cmo* mice (Fig. 2a). A HFD regimen induced remarkable changes in the colonic microbiota that was characterized by a suppression of disease-associated commensals (Fig. 2b, c). Most notably, LFD-fed *Pstpip2^cmo* mice displayed a time-dependent increase in *Prevotella* levels (Fig. 2d), which was significantly reduced in *Pstpip2^cmo* mice that were kept on a HFD (Fig. 2e). The latter group of HFD mice was further characterized by an expansion of *Lactobacillus* species in their intestinal tract (Fig. 2c). Diet-induced changes in the microbiota composition were not accompanied by readily detectable intestinal inflammation (Extended Data Fig. 3c–e). Moreover, we failed to detect bacteria in the peripheral organs of LFD-fed *Pstpip2^cmo* mice (Extended Data Fig. 3f). Together, these results show that inflammatory bone disease in *Pstpip2^cmo* mice is specifically characterized by an outgrowth of inflammation-associated intestinal commensals, which is suppressed by a HFD regimen.

We and others have previously shown that inflammatory bone disease in *Pstpip2^cmo* mice crucially relies on IL-1β (refs 5, 6). Given that *Pstpip2^cmo* mice on a HFD were markedly resistant to disease progression, we addressed whether HFD dampened IL-1β levels. *Pstpip2^cmo* mice that were fed a LFD had *Il1b* (which encodes the precursor protein pro-IL-1β) messenger RNA levels that were on average 60-fold higher than in footpads of healthy wild-type mice (Fig. 3a). In sharp contrast, HFD-fed *Pstpip2^cmo* mice had markedly suppressed local *Il1b* transcript levels that were comparable to those of healthy wild-type mice (Fig. 3a). In agreement with these observations, IL-1β protein concentrations were significantly increased in the footpads of LFD-fed *Pstpip2^cmo* mice, whereas those of HFD-fed *Pstpip2^cmo* mice were comparable to healthy controls (Fig. 3b and Extended Data Fig. 4a). Together, this suggests that HFD suppressed osteomyelitis in *Pstpip2^cmo* mice by dampening pro-IL-1β expression.


Figure 2 | Alterations in commensal microbiota landscape that are associated with *Pstpip2^cmo*-mediated osteomyelitic disease can be modified by changes in diet. a–c, Faecal samples were collected from wild-type, LFD *Pstpip2^cmo* and HFD *Pstpip2^cmo* mice at 10–12 weeks of age and 16S rRNA metagenomic sequencing was conducted. **a**, Heat map of fold differences in relative abundance of commensal bacteria. **b**, Principal coordinated analysis plot of faecal microbiota. **c**, Heat map of the top 20 commensal genera and species that differ between LFD *Pstpip2^cmo* and HFD *Pstpip2^cmo* mice are presented. **d**, *Prevotella* 16S rDNA copy numbers in wild-type and *Pstpip2^cmo* mice before (pre-disease: 3–6 weeks of age) and after (diseased: 10–16 weeks of age) the development of osteomyelitis. Each point represents an individual mouse, and the line represents the mean ± s.e.m. Data are representative of four independent experiments. **e**, 16S rDNA analysis of *Prevotella* abundance. Data are representative of four independent experiments. **P < 0.01, ***P < 0.001; Student's *t*-test.

Figure 3 | Microbiota-mediated regulation of IL-1β expression shapes inflammatory bone disease. a, Quantitative PCR with reverse transcription (qRT–PCR) analysis of relative *Il1b* expression in the footpads of 12–16-week-old wild-type, LFD *Pstpip2^cmo* and HFD *Pstpip2^cmo* mice. Each point represents an individual mouse, and the line represents the mean ± s.e.m. Combined data from three independent experiments. **b**, Protein levels of IL-1β in the hind paws. Combined data from two independent experiments. **c**, Relative *Il1b* mRNA expression levels in CD45+ cells isolated from the colons of specific-pathogen free (SPF) and germ-free (GF) WT mice. Two biological replicates, with two technical replicates each. **d, e**, *Pstpip2^cmo* mice were treated with a cocktail of broad-spectrum antibiotics in their drinking water (ABX). **d**, qRT–PCR analysis of colonic *Il1b* expression levels from 12–14-week-old *Pstpip2^cmo* mice that received either regular drinking water (*n* = 15) or antibiotics water (*n* = 9). **e**, Incidence of inflammatory bone disease. **f–i**, Young *Pstpip2^cmo* mice (3 weeks old) received PBS or faecal microbiota from diseased LFD *Pstpip2^cmo* or disease-free HFD *Pstpip2^cmo* mice by oral transplantation. **f**, 16S rDNA analysis of *Prevotella* copy numbers. **g**, Incidence of inflammatory bone disease. Combined data from three independent experiments. **h, i**, Representative footpad images (**h**) and haematoxylin and eosin micrographs (original magnification, ×2) (**i**). NS, not significant; *P < 0.05, **P < 0.01, ***P < 0.001; Student's *t*-test.

Given that HFD skewed the intestinal microbiota composition of *Pstpip2^cmo* mice (Fig. 2), we next asked whether the microbiota controlled *Il1b* expression. We found that *Il1b* levels in CD45⁺ cells that were isolated from the colons of germ-free wild-type mice were significantly lower than mice that were kept under specific pathogen-free conditions (Fig. 3c). Moreover, the levels of pro-IL-1β protein were considerably reduced in the hind paws of germ-free wild-type mice (Extended Data Fig. 4b). However, the expression of *Il1b* mRNA by CD45.2⁺ cells isolated from germ-free mice was greatly enhanced following *in vitro* stimulation with lipopolysaccharide (LPS), suggesting that these germ-free mice do not have any intrinsic defects in *Il1b* mRNA expression (Extended Data Fig. 4c). Notably, broad-spectrum antibiotics that significantly reduced *Prevotella* and *Flexispira* levels in LFD-fed *Pstpip2^cmo* mice (Extended Data Fig. 5a) also substantially decreased the levels of colonic *Il1b* in these mice (Fig. 3d). In addition, broad-spectrum antibiotics significantly protected LFD-fed *Pstpip2^cmo* mice from developing osteomyelitis (Fig. 3e). To address the role of the intestinal microbiota further, we performed faecal microbiota transplantation studies. Transplantation of the microbiota of diseased *Pstpip2^cmo* mice into wild-type mice failed to cause disease (Extended Data Fig. 5b). Similarly, LFD-fed *Pstpip2^cmo* mice also failed to transfer disease to co-housed wild-type and *Il1b*-deficient *Pstpip2^cmo* mice (Extended Data Fig. 5c, d). However, transplantation of the faecal microbiota of diseased (LFD-fed) *Pstpip2^cmo* mice to young LFD-fed *Pstpip2^cmo* mice by oral gavage promoted the expansion of *Prevotella* (Fig. 3f), and significantly accelerated the development of osteomyelitis relative to PBS-operated controls (Fig. 3g–i). Conversely, transplanting the microbiota of HFD-fed *Pstpip2^cmo* mice into

young LFD-fed *Pstpip2^cmo* mice greatly limited *Prevotella* outgrowth (Fig. 3f), and significantly protected mice from developing osteomyelitis (Fig. 3g–i). Although re-derivation of *Pstpip2^cmo* mice under germ-free conditions is needed to provide conclusive proof that commensal-derived factors are required to promote inflammatory bone disease, our findings clearly support the notion that diet-induced modulation of the microbiota composition regulates pro-IL-1β expression and osteomyelitis development in disease-susceptible *Pstpip2^cmo* mice.

The pro-IL-1β precursor protein is produced as a biologically inactive molecule that resides in the cytosol and needs to be proteolytically converted into mature IL-1β to gain biological activity. Caspase-1, a protease that is activated by inflammasome complexes, is the principal protease responsible for IL-1β maturation[11]. Neutrophil proteinase 3, elastase and caspase-8 were also recently shown to convert pro-IL-1β into its bioactive form[12–18]. Genetic deletion of caspase-1 and the related protease caspase-11 failed to rescue *Pstpip2^cmo* mice from inflammatory bone disease[5,6]. We therefore addressed the role of additional proteases in IL-1β-dependent osteomyelitis. To this end, *Pstpip2^cmo* mice were bred onto mice with gene-targeted deletions in neutrophil proteinase 3 and elastase (encoded by *Prtn3* and *Elane*, respectively). However, deletion of neither neutrophil proteinase 3 nor elastase rescued or delayed inflammatory bone disease in *Pstpip2^cmo* mice (Extended Data Fig. 6a, b). We next sought to examine the role of caspase-8 in *Pstpip2^cmo*-associated osteomyelitis. Mice deficient in caspase-8 are embryonic lethal[19–21], and this lethality is rescued by further deleting the necroptosis-regulating kinase RIPK3 (refs 22, 23). We thus bred *Casp8^−/−* *Ripk3^−/−* mice onto *Pstpip2^cmo* mice. Caspase-8 may act redundantly with caspase-1 in pro-IL-1β

Figure 4 | Compensatory processing of IL-1β by caspase-1 and caspase-8 in neutrophils drives inflammatory bone disease. a, Incidence of osteomyelitic disease in *Pstpip2^cmo*, *Pstpip2^cmo* × *Casp1^−/−*, *Pstpip2^cmo* × *Ripk3^−/−* × *Casp8^−/−* (double knockout, or DKO) *Pstpip2^cmo* × *Casp1^−/−* × *Ripk3^−/−* × *Casp8^−/−* (triple knockout, or TKO) and *Pstpip2^cmo* × *Il1b^−/−* mice over time. **b**, Representative haematoxylin and eosin staining of hind paw sections (original magnification, ×4). **c**, Western blot analysis of IL-1β regulation in the footpads of 10–12-week-old *Pstpip2^cmo*, *Pstpip2^cmo* × *Casp1^−/−*, *Pstpip2^cmo* × *Ripk3^−/−* × *Casp8^−/−*, *Pstpip2^cmo* × *Casp1^−/−* × *Ripk3^−/−* × *Casp8^−/−* and *Pstpip2^cmo* × *Il1b^−/−* mice. **d–f**, Wild-type and *Pstpip2^cmo* bone-marrow-derived macrophages (**d**) or neutrophils (**e, f**) were left untreated (UT) or were first primed with LPS for 3 h followed by stimulation with ATP (30 min) or silica (12 h). **d, e**, Secretion of IL-1β was measured by ELISA. Bar graphs depict mean ± s.e.m. One out of three biological replicates, with 2–3 technical replicates each. **f**, Western blot analysis of IL-1β in wild-type, *Pstpip2^cmo*, *Pstpip2^cmo* × *Casp1^−/−* and *Pstpip2^cmo* × *Casp1^−/−* × *Ripk3^−/−* × *Casp8^−/−* neutrophils. **g–i**, Wild-type and *Pstpip2^cmo* mice received either PBS or anti-Ly6G antibody every 4–5 days starting at 6 weeks of age. **g**, Incidence of inflammatory bone disease. **h, i**, Representative footpad images (**h**) and tail haematoxylin and eosin micrographs (original magnification, ×4) (**i**).
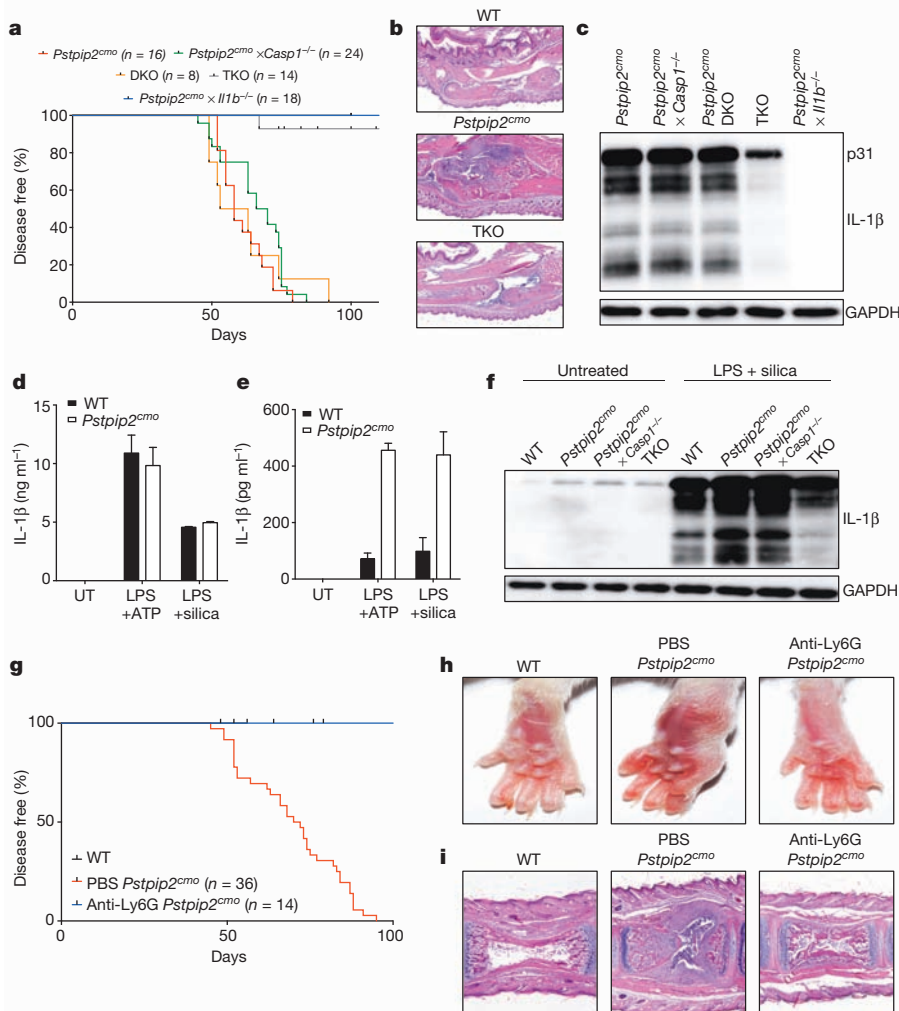
conversion under particular conditions[12–14,16], which we addressed by further deleting caspase-1 in *Casp8/Ripk3*-deficient *Pstpip2*[cmo] mice. As expected, *Pstpip2*[cmo] mice gradually developed inflammatory bone disease, with all mice being afflicted by 80 days (Fig. 4a). As reported[5,6], *Pstpip2*[cmo] mice lacking IL-1β were fully resistant to osteomyelitis development (Fig. 4a). *Ripk3*-deficient, *Casp1*-deficient and *Casp8/Ripk3*-deficient *Pstpip2*[cmo] mice developed osteomyelitis with similar kinetics to *Pstpip2*[cmo] mice (Fig. 4a and Extended Data Fig. 7a), which was also reflected in the extent of bone erosion and histopathology seen in these mice (Extended Data Fig. 7b, c). Notably, the combined deletion of caspase-1 and -8 provided significant protection against osteomyelitic disease (Fig. 4a, b). In agreement, pro-IL-1β expression levels were reduced and IL-1β maturation was virtually blunted in the footpads of *Pstpip2*[cmo] mice lacking both caspases (Fig. 4c and Extended Data Fig. 7d). In marked contrast, we observed spontaneous IL-1β maturation in footpads of *Pstpip2*[cmo] mice, as well as in mice lacking either caspase-1 or -8 (Fig. 4c).

Pstpip2[cmo] haematopoietic cells were recently shown to be sufficient to induce osteomyelitis in wild-type donor mice[5], suggesting that bone-marrow-derived cell populations are probably responsible for aberrant IL-1β production in *Pstpip2*[cmo] mice. We first evaluated the production of IL-1β by macrophages and neutrophils because these are the predominant immune cell types found in active osteoinflammatory lesions (Extended Data Fig. 1e). As reported[6], stimulation of LPS-primed *Pstpip2*[cmo] macrophages with NLRP3 inflammasome triggers such as ATP and silica triggered normal levels of secreted IL-1β (Fig. 4d and Extended Data Fig. 8a). In contrast, levels of IL-1β secreted by *Pstpip2*[cmo] neutrophils that were stimulated with these agents were at least fourfold higher than those of wild-type cells (Fig. 4e, f). Importantly, neutrophils of HFD-fed *Pstpip2*[cmo] mice expressed less pro-IL-1β (Extended Data Fig. 8b), and IL-1β maturation was markedly affected when compared to neutrophils of LFD-fed *Pstpip2*[cmo] mice (Extended Data Fig. 8c). By contrast, pro-IL-1β production and IL-1β maturation were not significantly different in macrophages of LFD- and HFD-fed *Pstpip2*[cmo] mice (Extended Data Fig. 8d). To ascertain the role of neutrophils in IL-1β-dependent osteomyelitis further, *Pstpip2*[cmo] mice were treated with anti-Ly6G antibodies to deplete neutrophils. Anti-Ly6G treatment led to marked reductions in circulating neutrophil counts (Extended Data Fig. 9a–c). Notably, neutrophil ablation conferred significant protection from clinical disease progression (Fig. 4g, h) and histopathological tissue damage (Fig. 4i).

Collectively, our findings presented here show that dietary intake determines the composition of the intestinal microbiota, and greatly influences disease outcome in osteomyelitis-susceptible *Pstpip2*[cmo] mice by upregulating pro-IL-1β levels. We further show that activation of caspases 1 and 8 in these mice result in spontaneous induction of IL-1β-driven neutrophilic osteomyelitis in *Pstpip2*[cmo] mice (Extended Data Fig. 10). These results suggest that diet-induced changes in the intestinal microbiota composition may promote autoinflammatory disease in susceptible individuals by increasing pro-IL-1β levels available for conversion by caspases 1 and 8.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 22 August 2014.
Published online 28 September 2014.

1. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2095–2128 (2012).
2. Ferguson, P. J. *et al.* A missense mutation in *pstpip2* is associated with the murine autoinflammatory disorder chronic multifocal osteomyelitis. *Bone* **38,** 41–47 (2006).
3. Chitu, V. *et al.* Primed innate immunity leads to autoinflammatory disease in PSTPIP2-deficient *cmo* mice. *Blood* **114,** 2497–2505 (2009).
4. Grosse, J. *et al.* Mutation of mouse *Mayp/Pstpip2* causes a macrophage autoinflammatory disease. *Blood* **107,** 3350–3358 (2006).
5. Lukens, J. R. *et al.* Critical role for inflammasome-independent IL-1β production in osteomyelitis. *Proc. Natl Acad. Sci. USA* **111,** 1066–1071 (2014).
6. Cassel, S. L. *et al.* Inflammasome-independent IL-1beta mediates autoinflammatory disease in Pstpip2-deficient mice. *Proc. Natl Acad. Sci. USA* **111,** 1072–1077 (2014).
7. Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12,** 5–9 (2011).
8. Montgomery, M. K. *et al.* Mouse strain-dependent variation in obesity and glucose homeostasis in response to high-fat feeding. *Diabetologia* **56,** 1129–1139 (2013).
9. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
10. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2014).
11. Dinarello, C. A. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood.* **117,** 3720–3732 (2011).
12. Maelfait, J. *et al.* Stimulation of Toll-like receptor 3 and 4 induces interleukin-1β maturation by caspase-8. *J. Exp. Med.* **205,** 1967–1973 (2008).
13. Gringhuis, S. I. *et al.* Dectin-1 is an extracellular pathogen sensor for the induction and processing of IL-1β via a noncanonical caspase-8 inflammasome. *Nature Immunol.* **13,** 246–254 (2012).
14. Antonopoulos, C., El Sanadi, C., Kaiser, W. J., Mocarski, E. S. & Dubyak, G. R. Proapoptotic chemotherapeutic drugs induce noncanonical processing and release of IL-1β via caspase-8 in dendritic cells. *J. Immunol.* **191,** 4789–4803 (2013).
15. Gurung, P. *et al.* FADD and caspase-8 mediate priming and activation of the canonical and noncanonical Nlrp3 inflammasomes. *J. Immunol.* **192,** 1835–1846 (2014).
16. Man, S. M. *et al. Salmonella* infection induces recruitment of Caspase-8 to the inflammasome to modulate IL-1β production. *J. Immunol.* **191,** 5239–5246 (2013).
17. Joosten, L. A. *et al.* Inflammatory arthritis in caspase 1 gene-deficient mice: contribution of proteinase 3 to caspase 1-independent production of bioactive interleukin-1β. *Arthritis Rheum.* **60,** 3651–3662 (2009).
18. Karmakar, M., Sun, Y., Hise, A. G., Rietsch, A. & Pearlman, E. Cutting edge: IL-1β processing during *Pseudomonas aeruginosa* infection is mediated by neutrophil serine proteases and is independent of NLRC4 and caspase-1. *J. Immunol.* **189,** 4231–4235 (2012).
19. Varfolomeev, E. E. *et al.* Targeted disruption of the mouse *Caspase 8* gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* **9,** 267–276 (1998).
20. Yeh, W. C. *et al.* FADD: essential for embryo development and signaling from some, but not all, inducers of apoptosis. *Science* **279,** 1954–1958 (1998).
21. Zhang, J., Cado, D., Chen, A., Kabra, N. H. & Winoto, A. Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1. *Nature* **392,** 296–300 (1998).
22. Kaiser, W. J. *et al.* RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* **471,** 368–372 (2011).
23. Oberst, A. *et al.* Catalytic activity of the caspase-8-FLIP$_L$ complex inhibits RIPK3-dependent necrosis. *Nature* **471,** 363–367 (2011).

Author Contributions J.R.L., M.L. and T.-D.K. designed the study; J.R.L., P.G., L.V.W., C.R.C. and S.R.B. performed experiments, and G.R.J. provided technical assistance. D.J.M. and R.A.C. analysed the 16S rRNA metagenomics data, and P.V. performed and analysed the histopathology data. J.R.L., M.L. and T.-D.K. analysed data and wrote the manuscript; T.-D.K. oversaw the project.

METHODS

Mice. *Pstpip2^{cmo}* (ref. 3), *Il1b$^{-/-}$* (ref. 24), *Casp1$^{-/-}$* (ref. 25), *Casp8$^{-/-}$* (ref. 26), *Ripk3$^{-/-}$* (ref. 27), *Elane$^{-/-}$* (ref. 28) and *Elane$^{-/-}$ Prtn3$^{-/-}$* (ref. 29) mice were previously described. *Pstpip2^{cmo}* were purchased from The Jackson Laboratory and are on the BALB/cJ background. All other mutant mice are on the C57BL/6 background. To generate the necessary controls and experimental mice for these experiments, mice that were heterozygous for both the *Pstpip2* and knockout allele(s) were used as breeders. Littermate controls were used to evaluate whether genetic deletions influence immune responses, IL-1β regulation and osteomyelitic disease development. Germ-free mice were obtained from Taconic. The number of mice per group used in an experiment is annotated in the corresponding figure legend as *n*. No gender differences were observed. *In vivo* experiments were controlled with age-matched littermates. The sample sizes were chosen to validate statistical analyses. All mice were kept in specific pathogen-free conditions within the Animal Resource Center at St Jude Children's Research Hospital. Animal studies were conducted under protocols approved by the Institutional Animal Care and Use Committee of St Jude Children's Research Hospital.

Diet. Feed that was high in fat and cholesterol was purchased from Research Diets Incorporated (stock number D12107) and consisted of 40% fat and 0.5% cholesterol. Standard low fat diet was obtained from LabDiet (stock number 5013) and consisted of 5% fat and 0% cholesterol.

Histopathology. Formalin-preserved paws and tails were processed and embedded in paraffin according to standard procedures. Haematoxylin and eosin (H&E) sections (5 μm) were examined by a pathologist blinded to the experimental groups. Tail and paw sections were scored based on the extent and severity of inflammation, pyogranulomatous, osteolysis and osteogenesis in a blinded fashion by a veterinary pathologist.

Micro-computed tomography. Micrographs of paws and tails fixed in formalin were made using an *ex vivo* micro-computed tomography scanner (LocusSP Specimen CT, GE Healthcare) at 28-μm isotropic voxel size, with 720 projections, an integration time of 1,700 ms, photon energy of 80 keV, and a current of 70 μA.

16S rRNA microbiome analysis. Fifty nanograms of purified DNA was prepared using Nextlex 16S v4 Amplicon-seq kit according to the manufacturer's instructions (Bioo Scientific). In brief, PCR primers targeted the fourth hypervariable domain of microbial 16S ribosomal RNA genes and simultaneously introduced sequences required for sequencing demultiplexing. Ampure XP PCR purification was used to clean up the PCR reactions (Beckman Coulter). PCR products were quantified using the Quant-iT PicoGreen assay (Invitrogen), normalized and pooled. Pooled samples were sequenced on a MiSeq sequencer (Illumina San Diego) according to manufacturer's instructions with modifications specified in the Nextflex 16S v4 kit. The 16S primers targeting the V4 region were aligned to the full set of sequences from the Greengenes database v13.5 using exonerate. Each sequence was truncated to include only the V4 region, the primer-matching regions, and an additional 40 bases on either side. Duplicate V4 regions were removed from the data set. All taxa labels from the removed duplicates were associated with the remaining representative V4 region sequence. Reads from each sample were aligned exhaustively to the non-redundant V4 sequences using USEARCH allowing a minimum sequence identity of 90%. All taxon labels associated with the top-scoring V4 region(s) were used to determine the taxon assignment of each read. The highest resolution non-conflicting taxon from all taxa associated with the top-scoring V4 region was assigned as the taxa for a read.

Relative proportions (*P*) of microbial taxa for each sample were assembled from the highest resolution sequence counts into a matrix with samples as columns and taxa as rows with proportions in cells. Columns were also assigned to a wild-type/knockout group according to the design. All relative proportions are transformed to near normality with a shifted logit-p transformation.

$$P_{\text{transformed}} = p^* = \frac{1}{2}\ln\left(\frac{p}{(1-p)}\right) + 7$$

Unpaired *t*-test with unequal variance for the normalized proportions (*P*$_{\text{transformed}}$) and a two-factor analysis of variance (ANOVA) model was used to investigate significant taxa. The transformed values are then normalized for each taxa to produce a signal-to-noise ratio (SNR).

$$\text{SNR} = \frac{p^* - \text{avg}(p^*)}{\text{stdev}(p^*)_{\text{pooled}}}; \quad \text{SNR} = \frac{\text{avg}(p^*)_{\text{WT}} - \text{avg}(p^*)_{\text{KO}}}{\text{stdev}(p^*)_{\text{pooled}}}$$

Signal-to-noise ratios are depicted in heat-map plots and principal component analyses generated with spotfire.

ELISA. Paw samples were snap frozen in liquid nitrogen and protein lysates were generated in RIPA lysis buffer supplemented with complete protease inhibitor cocktail (Roche) and PhosSTOP (Roche) using a tissue homogenizer. Debris was pelleted

and the supernatants were assessed by ELISA according to manufacturers' instructions (Milliplex and eBioscience).

Real-time PCR. Hind paw samples were snap frozen in liquid nitrogen and stored at −80 °C for later use. Tissue was homogenized in Trizol using a tissue homogenizer. Total RNA was isolated from the hind paws with Trizol (Invitrogen) according to the manufacturer's instructions. In brief, 200 μl of chloroform was added to the 1 ml of Trizol tissue lysate and the samples were incubated at room temperate for 5 min after vortexing. After centrifugation, the aqueous phase was transferred to a new tube and equal volumes of isopropanol were added. After incubation at room temperature for 10 min, the RNA was pelleted by centrifugation and then the RNA was washed twice in 70% ethanol before resuspension in ultrapure water. One microgram of RNA was reverse-transcribed to cDNA with random RNA-specific primers using the high-capacity cDNA reverse transcription kit (Applied Biosystems). Transcript levels of *Il1b*, *Cxcl1* (also known as *KC*), *Il6*, 16S *Prevotella*, 16S universal bacteria, *Actb* and *Gapdh* were analysed using SYBR-Green (Applied Biosystems) on an ABI7500 real-time PCR machine according to the manufacturers' recommendations. Relative expression was calculated using the ΔΔC_t standardization method.

Commensal bacteria depletion. Mice were treated with a broad-spectrum antibiotics regimen that contained 125 mg l^{-1} ciprofloxacin, 1 g l^{-1} bacitracin, 2 g l^{-1} streptomycin, 1.5 g l^{-1} metronidazole and 172 mg l^{-1} gentamycin in their drinking water.

***In vitro* macrophage stimulation.** Bone marrow-derived macrophages (BMDMs) were generated by culturing bone marrow cells in L-cell-conditioned IMDM medium supplemented with 10% FBS, 1% nonessential amino acid, and 1% penicillin-streptomycin for 5 days. BMDMs were seeded in 12-well cell culture plates and cultured overnight. To evaluate IL-1β production, BMDMs were primed with 2 μg ml^{-1} ultrapure *Escherichia coli*-derived LPS (Invivogen) for 3 h followed by 5 mM ATP (Sigma-Aldrich) for an additional 30 min. To measure IL-1β processing and production in response to stimulation with LPS and silica, cells were first primed with 2 μg ml^{-1} ultrapure *Escherichia coli*-derived LPS (Invivogen) for 3 h, and then were further activated with 500 μg ml^{-1} Min-U-Sil-5 silica (US Silica) for 5–12 h.

Neutrophil isolation and stimulation. Bone marrow cells were flushed from the femurs and tibias. Total bone marrow cells were passed through a 70-μm cell strainer and purified neutrophils were isolated from the interface of a 62.5% Percoll (GE Healthcare) gradient.

Western blotting. Hind paw protein lysates were collected in RIPA lysis buffer supplemented with complete protease inhibitor cocktail (Roche) and PhosSTOP (Roche) using a tissue homogenizer. Samples were clarified with at least two centrifugation steps to remove cellular debris. Lysates were resolved by SDS–PAGE and transferred to polyvinylidene difluoride (PVDF) membranes via electroblotting. Membranes were blocked in 5% non-fat milk and incubated overnight at 4 °C with primary antibodies. The following primary antibodies were used: anti-IL-1β clone D3H1Z (Cell Signaling) and anti-GAPDH (Cell Signaling). The membranes were probed with horseradish peroxidase (HRP)-tagged secondary antibodies at room temperature for 1 h. Immunoreactive proteins were visualized using the ECL method (Pierce).

Faecal transplantation. Fresh faecal samples were obtained from LFD-fed or HFD-fed *Pstpip2^{cmo}* mice and pellets were homogenized in PBS. Debris was pelleted by microcentrifugation and commensal bacteria were transplanted into young *Pstpip2^{cmo}* mice by oral transplantation every 2–4 days. Faecal reconstitution was confirmed by evaluating the intestinal abundance of *Prevotella* by 16S rDNA analysis in faecal microbiota transplantation mice 4–8 weeks later.

Neutrophil depletion. Wild-type and *Pstpip2^{cmo}* mice received either PBS or 500 μg per mouse anti-Ly6G antibody (clone IA8) by intraperitoneal injection every 4–5 days starting at 6 weeks of age and the incidence of inflammatory bone disease was evaluated over time. Depletion of neutrophils was confirmed by FACS staining for CD45.2$^+$ CD11b$^+$ Gr-1$^+$ cells in the peripheral blood.

Statistical analysis. All results are presented as mean ± standard error. We performed statistical analysis using the two-tailed Student's *t*-test. *P* values are denoted by *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

24. Shornick, L. P. *et al.* Mice deficient in IL-1β manifest impaired contact hypersensitivity to trinitrochlorobenzene. *J. Exp. Med.* **183,** 1427–1436 (1996).
25. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440,** 233–236 (2006).
26. Salmena, L. *et al.* Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. *Genes Dev.* **17,** 883–895 (2003).
27. Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF-κBs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell. Biol.* **24,** 1464–1469 (2004).
28. Belaaouaj, A. *et al.* Mice lacking neutrophil elastase reveal impaired host defense against gram negative bacterial sepsis. *Nature Med.* **4,** 615–618 (1998).
29. Kessenbrock, K. *et al.* Proteinase 3 and neutrophil elastase enhance inflammation in mice by inactivating antiinflammatory progranulin. *J. Clin. Invest.* **118,** 2438–2447 (2008).
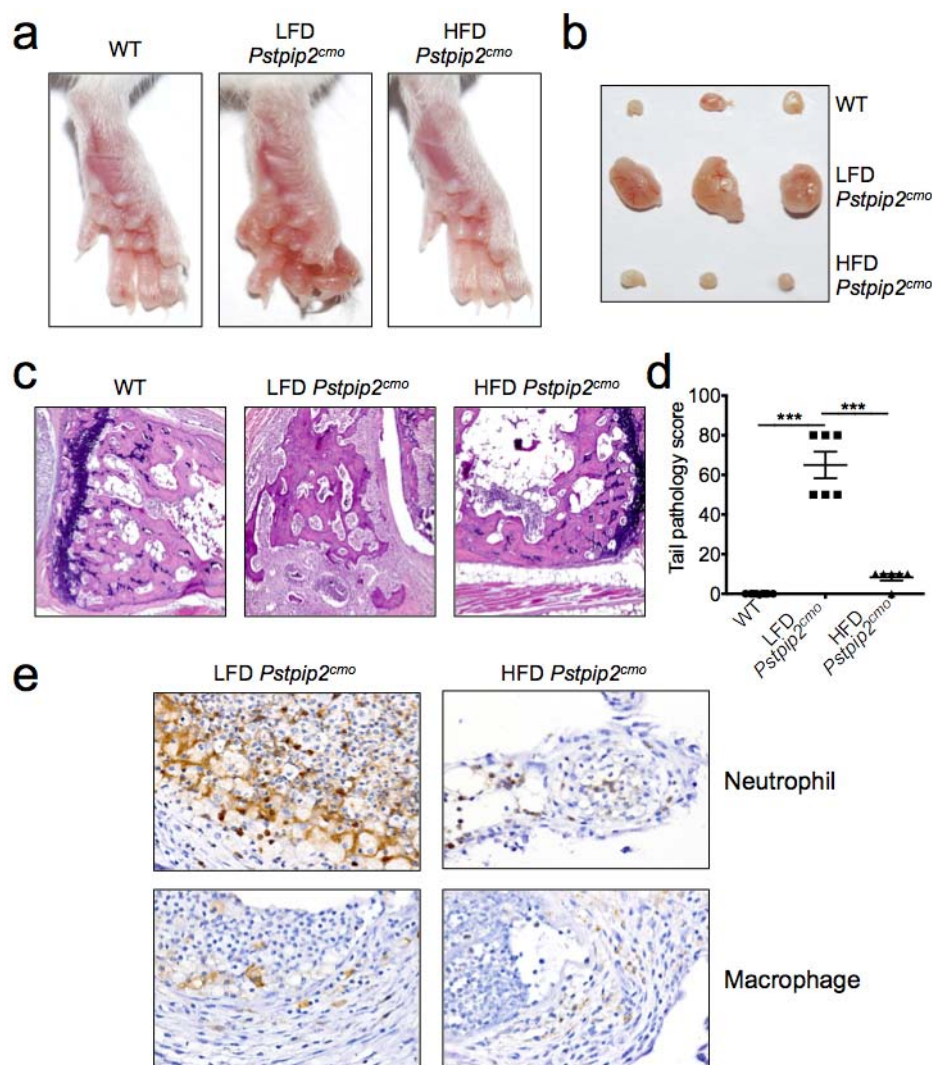
Extended Data Figure 1 | Placing *Pstpip2^cmo* mice on a high-fat and cholesterol diet limits the development of inflammatory bone disease. **a–e**, Wild-type and *Pstpip2^cmo* mutant mice were fed a LFD or HFD. Representative hind paw images (**a**) and representative pictures of popliteal lymph nodes (**b**) from wild-type, LFD *Pstpip2^cmo* and HFD *Pstpip2^cmo* mice at 12–14 weeks of age. **c**, **d**, Haematoxylin and eosin staining (original magnification, ×20) (**c**) and pathology scores (**d**) of tail samples from 12–14-week-old wild-type, LFD *Pstpip2^cmo* and HFD *Pstpip2^cmo* mice. Pathology scores were assigned in a blinded fashion by a veterinary pathologist based on the extent and severity of inflammation, osteolysis and osteogenesis. **e**, Representative immunostaining of neutrophils and macrophages in hind paw sections from 14–18-week-old *Pstpip2^cmo* mice that were fed either a LFD or a HFD (original magnification, ×60). ***$P < 0.001$; Student's *t*-test.
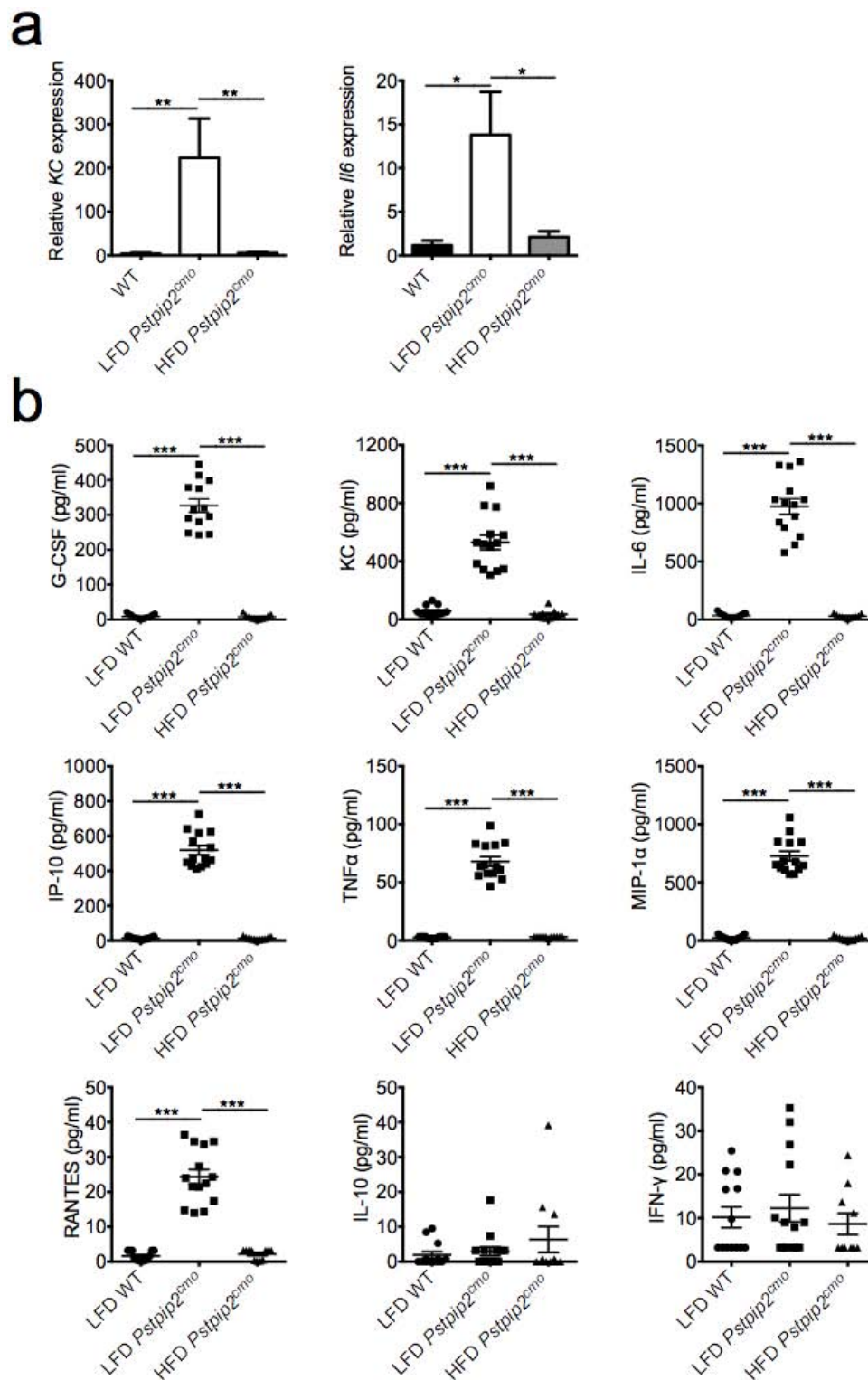
Extended Data Figure 2 | Consumption of a HFD limits hyperinflammatory cytokine production in *Pstpip2^cmo* mutant mice.
a, Wild-type and *Pstpip2^cmo* mutant mice were fed a LFD or HFD for 12 weeks. Relative expression of *Cxcl1* (wild type $n = 8$; LFD *Pstpip2^cmo* $n = 4$; HFD *Pstpip2^cmo* $n = 9$) and *Il6* (wild type $n = 11$; LFD *Pstpip2^cmo* $n = 10$; HFD *Pstpip2^cmo* $n = 8$) in the hind paws was determined by qRT–PCR. The bar graphs depict combined data from two independent experiments. Data are shown as mean ± s.e.m. **b**, Wild-type and *Pstpip2^cmo* mutant mice were fed a LFD or a HFD for 12 weeks and cytokines levels in the hind paws were determined by ELISA. Combined data are from two independent experiments. Each point represents an individual mouse, and the line represents the mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; Student's *t*-test.
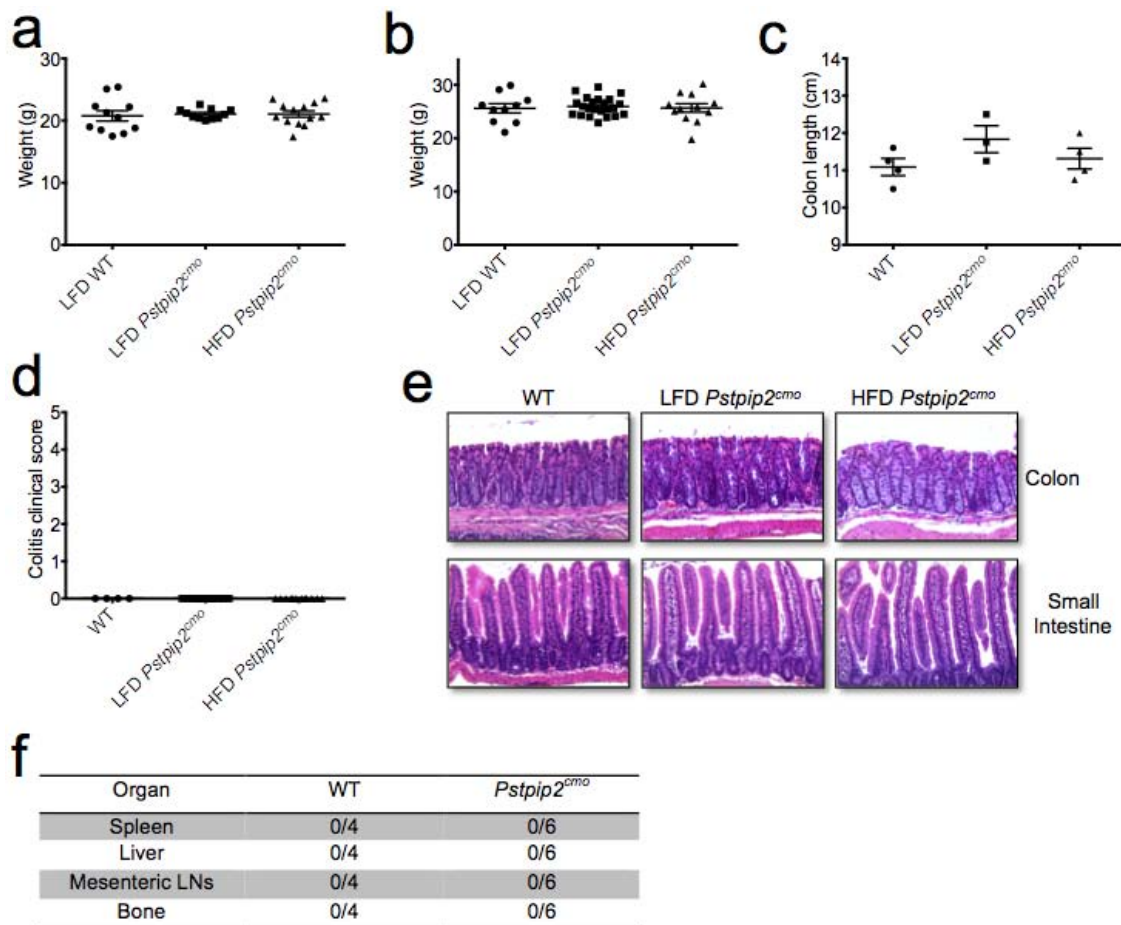
Extended Data Figure 3 | Placing *Pstpip2^cmo* mice on a HFD does not cause abnormal weight gain, intestinal inflammation or extraintestinal translocation of commensal bacteria. a, b, Wild-type BALB/cJ and *Pstpip2^cmo* mice were fed *ad libidum* a LFD or a HFD. Body weight was measured in age-matched female (a) and male (b) mice at 12–16 weeks of age. Each point represents an individual mouse and the line represents the mean ± s.e.m. Data were combined from three independent experiments. c–e, Colon length

(c), colitis score based on rectal bleeding and stool consistency (d) and representative haematoxylin and eosin-stained sections (original magnification, ×20) (e) of the intestinal tract of LFD- and HFD-fed *Pstpip2^cmo* mice aged 14–18 weeks. f, Presence of commensal bacteria in the spleen, liver, mesenteric lymph nodes and bone of wild-type and diseased LFD-fed *Pstpip2^cmo* mice was evaluated by Gram staining and 16S rDNA qPCR analysis of eubacteria.

a

WT | LFD *Pstpip2cmo* | HFD *Pstpip2cmo*

IL-1β

GAPDH

b

SPF WT | GF WT | *Pstpip2cmo* × *Il1b−/−*

proIL-1β

GAPDH

c
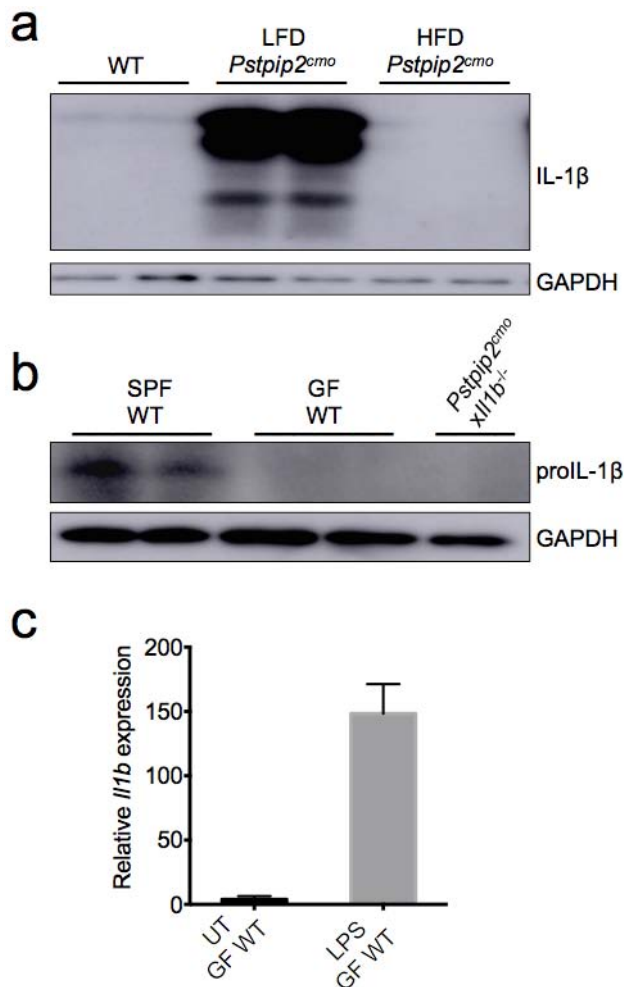
Extended Data Figure 4 | Dietary- and microbiota-associated factors influence the production of pro-IL-1β. a, Footpad homogenates of 12–16-week-old wild-type, LFD-fed *Pstpip2cmo* and HFD-fed *Pstpip2cmo* mice were immunoblotted for IL-1β. Data are representative of three independent experiments. **b,** Footpads samples were collected from 10–14-week-old specific pathogen-free wild-type, germ-free wild-type and *Pstpip2cmo* × *Il1b−/−* mice and pro-IL-1β protein levels were determined by western blotting. **c,** CD45+ cells were isolated from the colons of germ-free wild-type mice and cells were left untreated or stimulated with LPS for 1 h. Relative *Il1b* mRNA expression levels were determined by qRT–PCR. Two biological replicates, with two technical replicates each.
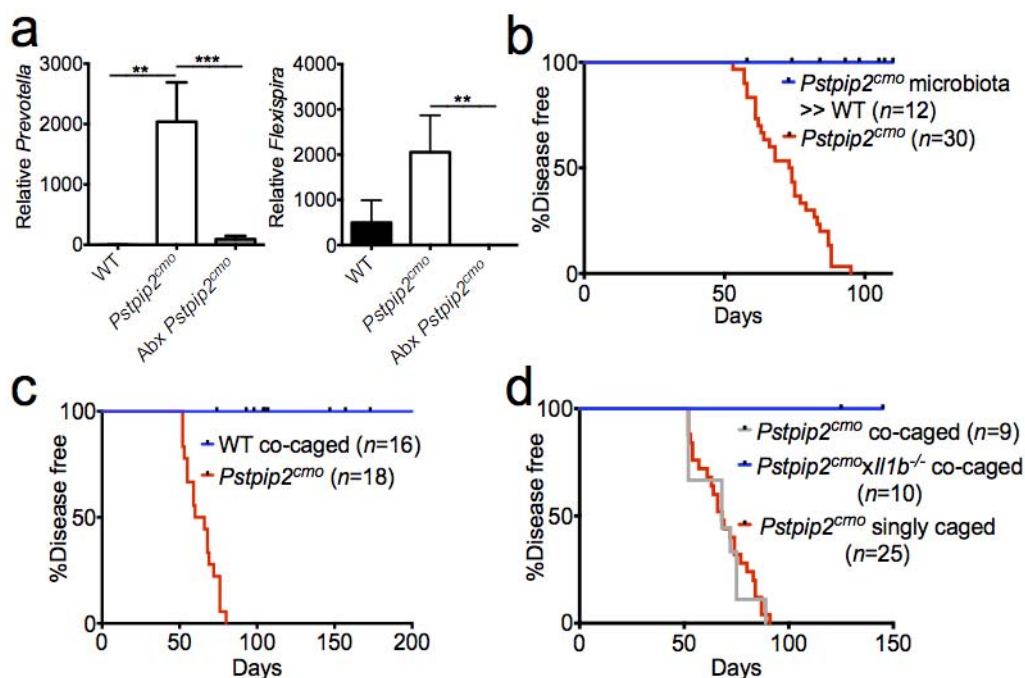
Extended Data Figure 5 | Co-housing does not alter disease progression in LFD-fed *Pstpip2cmo* mice. a, *Pstpip2cmo* mice were treated with a cocktail of broad-spectrum antibiotics in their drinking water. Faecal samples were collected from wild-type (*n* = 5) and *Pstpip2cmo* mice that received either regular drinking water (*n* = 5) or antibiotics water (*n* = 11) 5–7 weeks later. *Prevotella* and *Flexispira* 16S rDNA copy numbers were quantified and normalized to total bacteria. The bar graphs depict the mean ± s.e.m. **b**, Faecal microbiota from diseased *Pstpip2cmo* mice was orally transplanted into wild-type mice (*Pstpip2cmo* microbiota ≫ wild type) and the incidence of inflammatory bone disease in control *Pstpip2cmo* and faecal transplantation mice was evaluated. **c, d**, *Pstpip2cmo* mice were singly housed or co-housed with wild-type (**c**) or *Il1b*-deficient *Pstpip2cmo* (**d**) mice. Clinical development of bone deformity and arthritic inflammation in hind paws and tails was monitored over time. **P < 0.01, ***P < 0.001; Student's *t*-test.
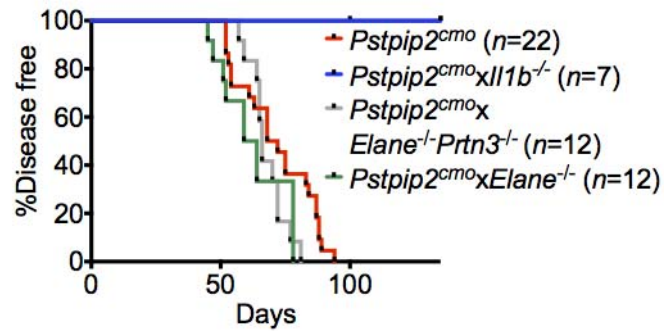
Extended Data Figure 6 | The neutrophil associated proteases elastase and proteinase 3 are not required for *Pstpip2^{cmo}*-mediated bone disease.
a, Incidence of inflammatory bone disease in *Pstpip2^{cmo}*, *Pstpip2^{cmo}* × *Elane^{−/−}*, *Pstpip2^{cmo}* × *Elane^{−/−} Prtn3^{−/−}* and *Pstpip2^{cmo}* × *Il1b^{−/−}* mice. **b**, Representative footpad images from wild-type, *Pstpip2^{cmo}*, *Pstpip2^{cmo}* × *Elane^{−/−}*, *Pstpip2^{cmo}* × *Elane^{−/−} Prtn3^{−/−}* and *Pstpip2^{cmo}* × *Il1b^{−/−}* mice.
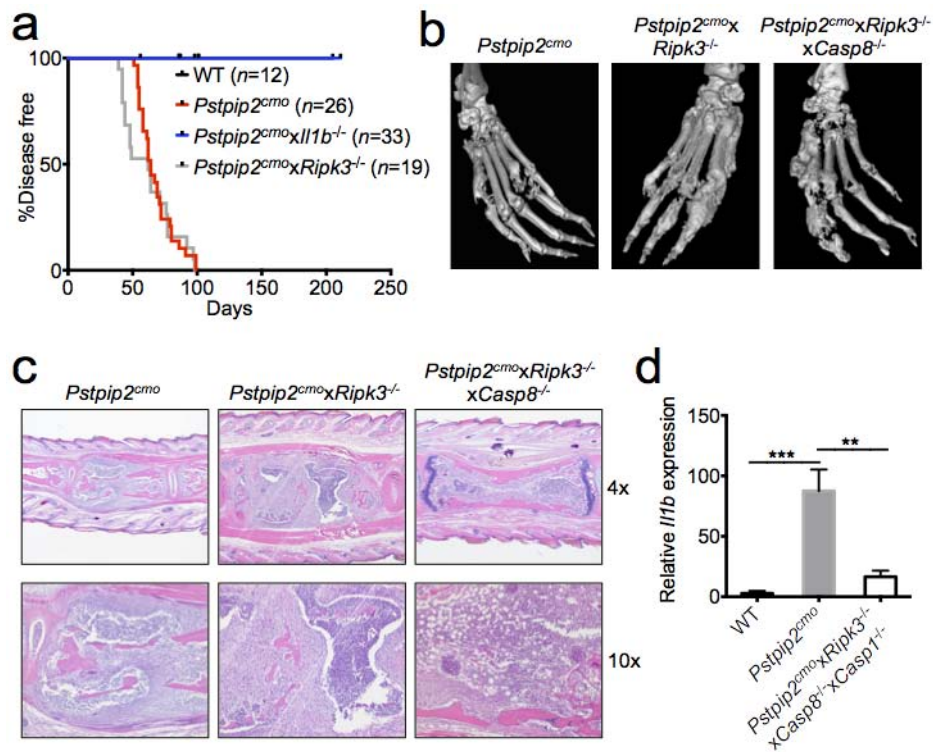
Extended Data Figure 7 | Combined deletion of RIPK3 and caspase-8 does not provide protection against *Pstpip2^cmo*-mediated osteomyelitis.
a, Incidence of osteomyelitic bone disease in wild-type, *Pstpip2^cmo*, *Pstpip2^cmo* × *Il1b*^−/− and *Pstpip2^cmo* × *Ripk3*^−/− mice. **b**, Representative isosurface micro-computed tomography images of hind paw samples from 12–18-week-old *Pstpip2^cmo*, *Pstpip2^cmo* × *Ripk3*^−/− and *Pstpip2^cmo* × *Ripk3*^−/− × *Casp8*^−/− mice. **c**, Representative haematoxylin and eosin-stained sections of inflammatory caudal vertebrae bone lesions in diseased *Pstpip2^cmo*, *Pstpip2^cmo* × *Ripk3*^−/− and *Pstpip2^cmo* × *Ripk3*^−/− × *Casp8*^−/− mice (original magnification, ×4 (top) and ×10 (bottom)). **d**, qRT–PCR analysis of *Il1b* expression in footpads of wild-type (*n* = 7), *Pstpip2^cmo* (*n* = 7) and *Pstpip2^cmo* × *Ripk3*^−/− × *Casp8*^−/− × *Casp1*^−/− (*n* = 7) mice aged 12–16 weeks. Data are expressed as mean ± s.e.m. of combined data from two independent experiments. **P < 0.01, ***P < 0.001; Student's *t*-test.
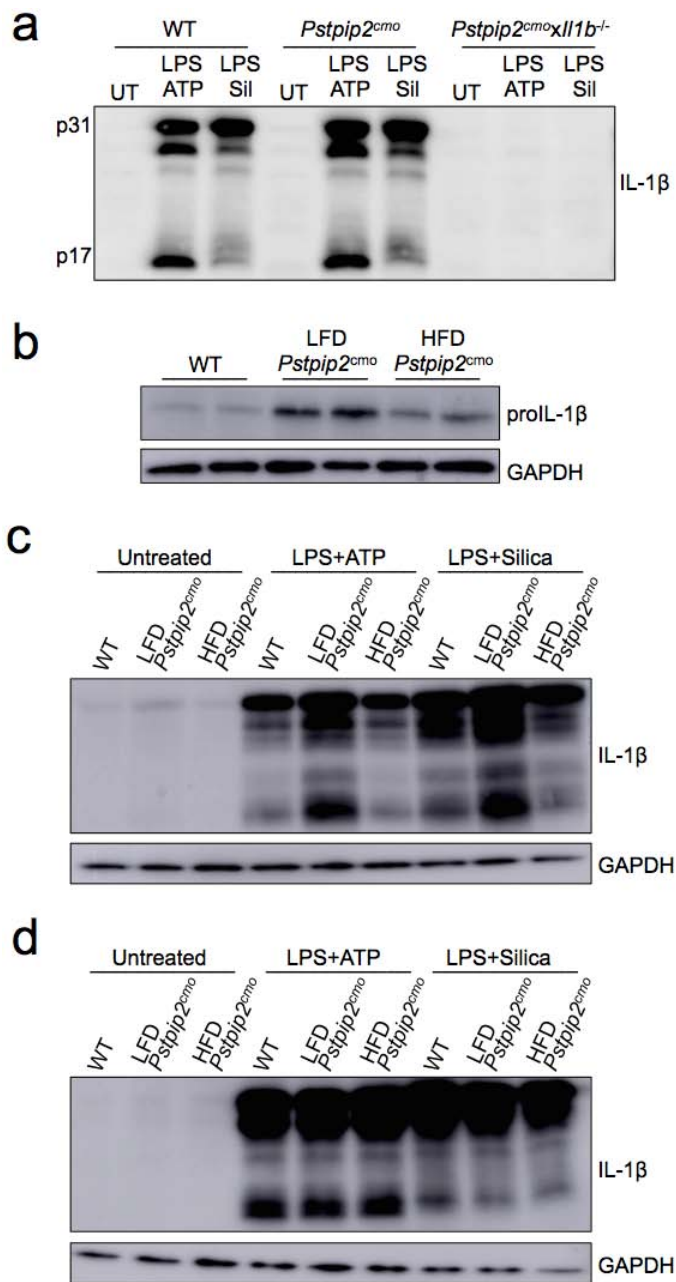
Extended Data Figure 8 | Reduced pro-IL-1β expression and IL-1β maturation in neutrophils isolated from HFD-fed *Pstpip2^cmo^* mice.
a, Wild-type, *Pstpip2^cmo^* and *Pstpip2^cmo^* × *Il1b^−/−^* bone-marrow-derived macrophages were left untreated or were primed with LPS for 3 h followed by stimulation with ATP (30 min) or silica (12 h), and IL-1β processing was evaluated by western blot. Data are representative of three independent experiments. **b**, Western blotting for pro-IL-1β in untreated neutrophils that were purified from wild-type, LFD-fed *Pstpip2^cmo^* and HFD-fed *Pstpip2^cmo^* mice. Data are representative of two independent experiments. **c**, **d**, Neutrophils (**c**) or macrophages (**d**) from wild-type, LFD-fed *Pstpip2^cmo^* and HFD-fed *Pstpip2^cmo^* mice were left untreated, or primed with LPS for 3 h and then stimulated with ATP (30 min) or silica (12 h), and IL-1β processing was evaluated by western blotting. Data are representative of two independent experiments.
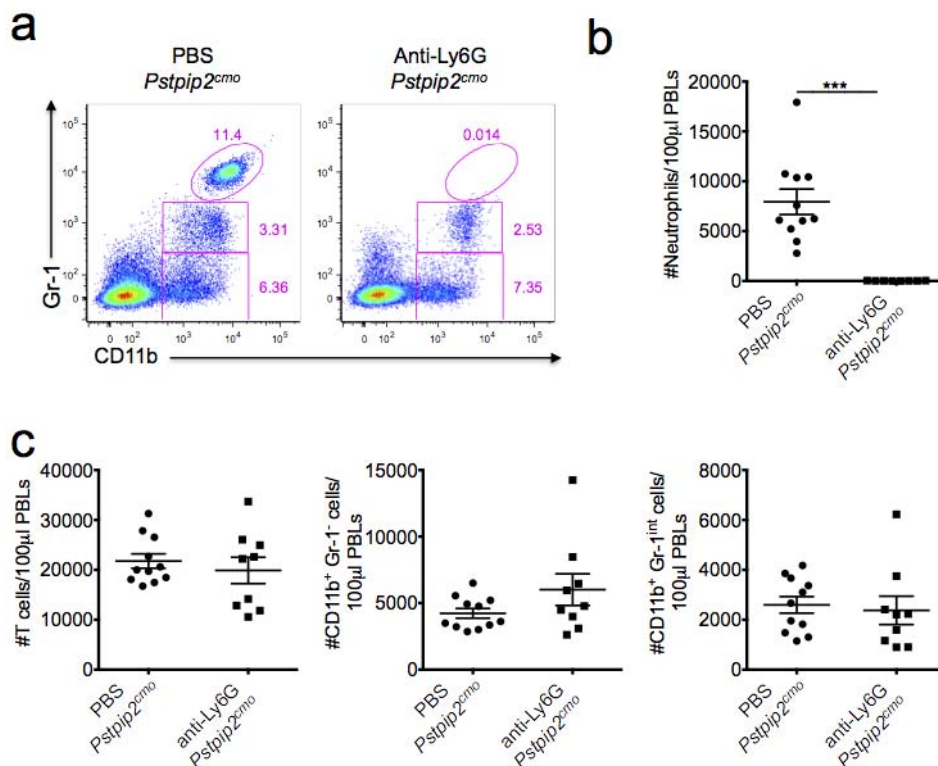
Extended Data Figure 9 | Depletion of neutrophils in anti-Ly6G treated *Pstpip2cmo* mutant mice. Wild-type and *Pstpip2cmo* mice received either PBS or 500 μg per mouse anti-Ly6G antibody (clone IA8) by intraperitoneal injection every 4–5 days starting at 6 weeks of age. **a–c,** Two weeks after the first anti-Ly6G treatment, FACS analysis was performed on peripheral blood leukocytes (PBLs). **a,** Representative FACS plots of Gr-1 and CD11b expression on CD45.2$^+$ gated cells. **b,** Enumeration of CD45.2$^+$ Gr-1^{hi} CD11b$^+$ neutrophils in equal volumes of peripheral blood. **c,** Numbers of T cells (CD45.2$^+$ TCRβ$^+$), CD45.2$^+$ Gr-1$^-$ CD11b$^+$ monocytes/macrophages and CD45.2$^+$ Gr-1^{int} CD11b$^+$ cells in equal volumes of peripheral blood. Each point represents an individual mouse and the line represents the mean ± s.e.m. ***$P < 0.001$; Student's t-test.
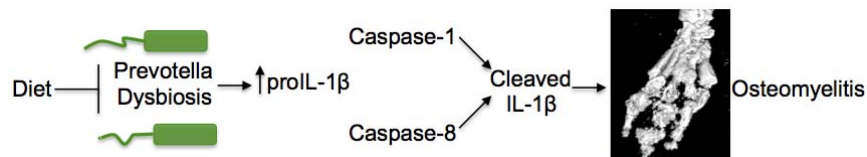
Extended Data Figure 10 | Dietary modulation of the intestinal microbiota composition drives autoinflammatory osteomyelitis by setting pro-IL-1β levels available for maturation by caspases 1 and 8. Proposed model highlighting how dysbiosis and processing of IL-1β by caspases 1 and 8 contribute to the development of inflammatory bone disease.

LETTER

Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG

Parveen Goyal[1,2], Petya V. Krasteva[3,4], Nani Van Gerven[1,2], Francesca Gubellini[3,4], Imke Van den Broeck[1,2], Anastassia Troupiotis-Tsaïlaki[5], Wim Jonckheere[1,2], Gérard Péhau-Arnaudet[4], Jerome S. Pinkner[6], Matthew R. Chapman[7], Scott J. Hultgren[6], Stefan Howorka[8], Rémi Fronzes[3,4] & Han Remaut[1,2]

Curli are functional amyloid fibres that constitute the major protein component of the extracellular matrix in pellicle biofilms formed by Bacteroidetes and Proteobacteria (predominantly of the α and γ classes)[1–3]. They provide a fitness advantage in pathogenic strains and induce a strong pro-inflammatory response during bacteraemia[1,4,5]. Curli formation requires a dedicated protein secretion machinery comprising the outer membrane lipoprotein CsgG and two soluble accessory proteins, CsgE and CsgF[6,7]. Here we report the X-ray structure of *Escherichia coli* CsgG in a non-lipidated, soluble form as well as in its native membrane-extracted conformation. CsgG forms an oligomeric transport complex composed of nine anticodon-binding-domain-like units that give rise to a 36-stranded β-barrel that traverses the bilayer and is connected to a cage-like vestibule in the periplasm. The transmembrane and periplasmic domains are separated by a 0.9-nm channel constriction composed of three stacked concentric phenylalanine, asparagine and tyrosine rings that may guide the extended polypeptide substrate through the secretion pore. The specificity factor CsgE forms a nonameric adaptor that binds and closes off the periplasmic face of the secretion channel, creating a 24,000 Å³ pre-constriction chamber. Our structural, functional and electrophysiological analyses imply that CsgG is an ungated, non-selective protein secretion channel that is expected to employ a diffusion-based, entropy-driven transport mechanism.

Curli are bacterial surface appendages that have structural and physical characteristics of amyloid fibrils, best known from human degenerative diseases[7–9]. However, the role of bacterial amyloids such as curli are to facilitate biofilm formation[4,10]. Unlike pathogenic amyloids, which are the product of protein misfolding, curli formation is coordinated by proteins encoded in two dedicated operons, *csgBAC* (*curli specific genes BAC*) and *csgDEFG* in *Escherichia coli* (Extended Data Fig. 1)[6,7]. After secretion, CsgB nucleates CsgA subunits into curli fibres[7,11,12]. Secretion and extracellular deposition of CsgA and CsgB are dependent on two soluble accessory factors, respectively CsgE and CsgF, as well as CsgG, a 262-residue lipoprotein located in the outer membrane[13–16]. Because of the lack of hydrolysable energy sources or ion gradients at the outer membrane, CsgG falls into a specialized class of protein translocators that must operate through an alternatively energized transport mechanism. In the absence of a structural model, the dynamic workings of how CsgG promotes the secretion and assembly of highly stable amyloid-like fibres in a regulated fashion across a biological membrane has so far remained enigmatic.

Before insertion into the outer membrane, lipoproteins are piloted across the periplasm by means of the lipoprotein localization (Lol) pathway[17]. We observed that non-lipidated CsgG (CsgG$_{C1S}$) could be isolated as a soluble periplasmic intermediate, analogous to the pre-pore forms observed in pore-forming proteins and toxins[18]. CsgG$_{C1S}$ was found predominantly as monomers, in addition to a minor fraction of discrete oligomeric

complexes (Extended Data Fig. 2)[19]. The soluble CsgG$_{C1S}$ oligomers were crystallized and their structure was determined to 2.8 Å, revealing a hexadecameric particle with eight-fold dihedral symmetry (D_8), consisting of two ring-shaped octameric complexes (C_8) that are joined in a tail-to-tail interaction (Extended Data Fig. 2 and Fig. 1). The CsgG$_{C1S}$ protomer shows an anticodon-binding domain (ABD)-like fold that is extended with two α-helices at the amino and carboxy termini (αN and αC, respectively; Fig. 1 and Extended Data Fig. 3a–c). Additional CsgG-specific elements are an extended loop linking β1 and α1, two insertions in the loops connecting β3–β4 and β5–α3, and an extended α2 helix that is implicated in CsgG oligomerization by packing between adjacent monomers (Fig. 1b). Further inter-protomer contacts are formed between the back of the β3–β5 sheet and the extended β1–α1 loop (Extended Data Fig. 3d, e).

In the CsgG$_{C1S}$ structure, residues 1–17, which would link α1 to the N-terminal lipid anchor, are disordered and no obvious transmembrane (TM) domain can be discerned (Fig. 1). Attenuated total reflection Fourier transform infrared spectroscopy (ATR–FTIR) of CsgG$_{C1S}$ and native, membrane-extracted CsgG revealed that the latter has a higher absorption in the β-sheet region (1,625–1,630 cm^{-1}) and a concomitant reduction in the random coil and α-helical regions (1,645–1,650 cm^{-1} and 1,656 cm^{-1}, respectively; Fig. 2a), suggesting that membrane-associated CsgG contains a β-barrel domain. Candidate sequence stretches for β-strand formation are found in the poorly ordered, extended loops connecting β3–β4 (residues 134–154) and β5–α3 (residues 184–204); deletion
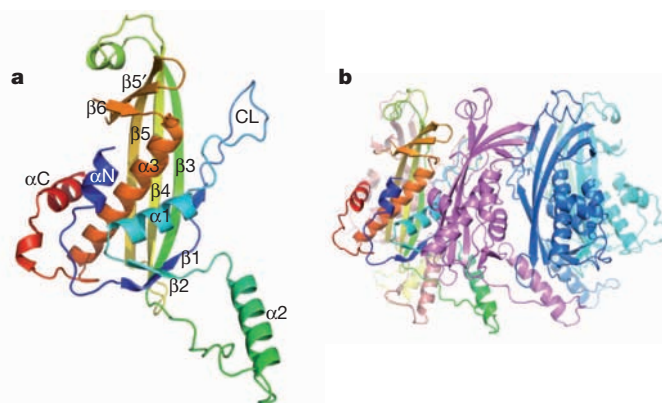
Figure 1 | X-ray structure of CsgG$_{C1S}$ in pre-pore conformation. a, Ribbon diagram of the CsgG$_{C1S}$ monomer coloured as a blue-to-red rainbow from N terminus to C terminus. Secondary structure elements are labelled according to the ABD-like fold, with the additional N-terminal and C-terminal α-helices and the extended loop connecting β1 and α1 labelled αN, αC and C-loop (CL), respectively. **b**, Side view of the CsgG$_{C1S}$ C_8 octamer with subunits differentiated by colour and one subunit oriented and coloured as in **a**.

[1]Structural and Molecular Microbiology, Structural Biology Research Center, VIB, Pleinlaan 2, 1050 Brussels, Belgium. [2]Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [3]Unité G5 Biologie structurale de la sécrétion bactérienne, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. [4]UMR 3528, CNRS, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. [5]Structure et Fonction des Membranes Biologiques (SFMB), Université Libre de Bruxelles, 1050 Brussels, Belgium. [6]Department of Molecular Microbiology and Microbial Pathogenesis, Washington University in Saint Louis School of Medicine, St Louis, Missouri 63110-1010, USA. [7]Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA. [8]Department of Chemistry, Institute for Structural and Molecular Biology, University College London, London WC1H 0AJ, UK.
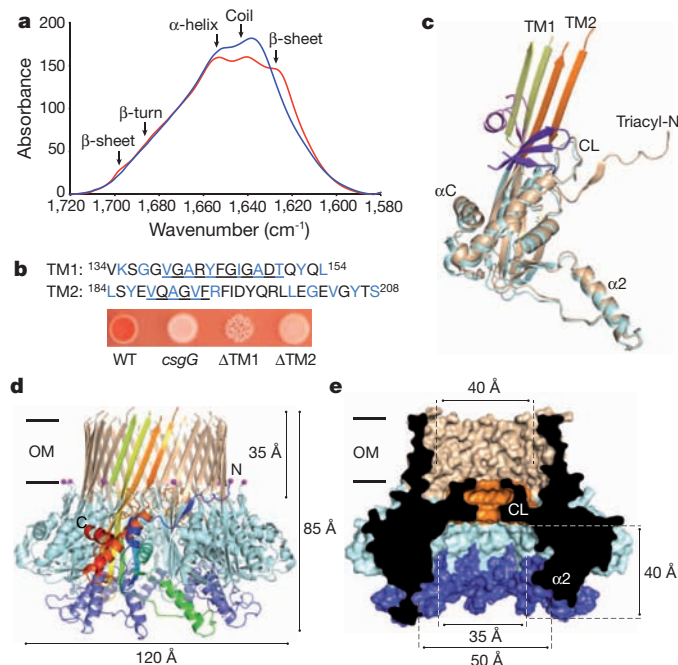
Figure 2 | Structure of CsgG in its channel conformation. a, Amide I region (1,700–1,600 cm^{-1}) of ATR–FTIR spectra of CsgG$_{C1S}$ (blue) and membrane-extracted CsgG (red). **b**, TM1 and TM2 sequence (bilayer-facing residues in blue) and Congo red binding of *E. coli* BW25141Δ*csgG* complemented with wild-type *csgG* (WT), empty vector or *csgG* lacking the underlined fragments of TM1 or TM2. Data are representative of three biological replicates. **c**, Overlay of CsgG monomer in pre-pore (light blue; TM1 pink, TM2 purple) and channel conformation (tan; TM1 green, TM2 orange). CL, C-loop. **d**, **e**, Side view (**d**) and cross-sectional view (**e**) of CsgG nonamers in ribbon and surface representation; helix 2, the core domain and TM hairpins are shown in blue, light blue and tan, respectively. A single protomer is coloured as in Fig. 1a. Magenta spheres show the position of Leu 2. OM, outer membrane.

of these resulted in the loss of curli formation (Fig. 2b). The crystal structure of detergent-extracted CsgG confirmed a conformational rearrangement of both regions into two adjacent β-hairpins, extending the β-sheet formed by β3–β4 (TM1) and β5–α3 (TM2) (Fig. 2c). Their juxtaposition in the CsgG oligomer gave rise to a composite 36-stranded β-barrel (Fig. 2d). Whereas the crystallized CsgG$_{C1S}$ oligomers showed a D_8 symmetry, the CsgG structure showed D_9 symmetry, with CsgG protomers retaining equivalent interprotomer contacts, except for a 5° rotation relative to the central axis and a 4 Å translation along the radial axes (Extended Data Fig. 2). This observation is reconciled in the in-solution oligomeric states revealed by single-particle electron microscopy, which exclusively found C_9 and D_9 symmetries for membrane-extracted CsgG (Extended Data Fig. 2). The predominant presence of monomers in the non-lipidated sample and the symmetry mismatch with the membrane-bound protein argue that before membrane insertion, CsgG is targeted to the outer membrane in a monomeric, LolA-bound form and that the C_8 and D_8 particles are an artefact of highly concentrated solutions of CsgG$_{C1S}$. Furthermore, we show that the C_9 nonamer rather than the D_9 complex forms the physiologically relevant particle, because in isolated *E. coli* outer membranes, cysteine substitutions in residues enclosed by the observed tail-to-tail dimerization are accessible to labelling with maleimide-polyethylene glycol (PEG, 5 kDa; Extended Data Fig. 4).

Thus, CsgG forms a nonameric transport complex 120 Å in width and 85 Å in height. The complex traverses the outer membrane through a 36-stranded β-barrel with an inner diameter of 40 Å (Fig. 2e). The N-terminal lipid anchor is separated from the core domain by an 18-residue linker that wraps over the adjacent protomer (Extended Data Fig. 3d). The diacylglycerol- and amide-linked acyl chain on the N-terminal Cys are not resolved in the electron density maps, but on the basis of the location of Leu 2 the lipid anchor is expected to flank the outer wall of the

β-barrel. On the periplasmic side, the transporter forms a large solvent–accessible cavity with an inner diameter of 35 Å and a height of 40 Å that opens to the periplasm in a 50 Å mouth formed by helix 2 (Fig. 2e). At its apex, this periplasmic vestibule is separated from the TM channel by a conserved 12-residue loop connecting β1 to α1 (C-loop; Figs 2e and 3a, b), which constricts the secretion conduit to a solvent-excluded diameter of 9.0 Å (Fig. 3a, c). These pore dimensions would be compatible with the residence of one or two (for example a looped structure) extended polypeptide segments, with five residues spanning the height of the constriction (Extended Data Fig. 5). The luminal lining of the constriction is composed of three stacked concentric rings formed by the side chains of residues Tyr 51, Asn 55 and Phe 56 (Fig. 3a, b). In the anthrax PA$_{63}$ toxin, a topologically equivalent concentric Phe ring (referred to as a φ-clamp) lines the entry of the translocation channel and catalyses polypeptide capture and passage[20–22]. Multiple sequence alignment of CsgG-like translocators shows the absolute conservation of Phe 56 and the conservative variation of Asn 55 to Ser or Thr (Extended Data Fig. 6). Mutation of Phe 56 or Asn 55 to Ala leads to a near loss of curli production (Fig. 3d), whereas a Asn 55→Ser substitution retains wild-type secretion levels, together alluding to the requirement of the stacked configuration of a φ-clamp followed by a hydrogen-bond donor/acceptor in the CsgG constriction (Fig. 3b and Extended Data Fig. 6).

Single-channel current recordings of CsgG reconstituted in planar phospholipid bilayers led to a steady current of 43.1 ± 4.5 pA ($n = 33$) or −45.1 ± 4.0 pA ($n = 13$) using standard electrolyte conditions and a potential of +50 mV or −50 mV, respectively (Fig. 3e, f and Extended Data Fig. 7). The observed current was in good agreement with the predicted value of 47 pA calculated on the basis of a simple three-segment pore model and the dimensions of the channel lumen seen in the X-ray structure (Fig. 3c). A second, low-conductance conformation can also be observed under negative electrical field potential (−26.2 ± 3.6 pA ($n = 13$); Extended Data Fig. 7). It is unclear, however, whether this species is present under physiological conditions.

Our structural data and single-channel recordings imply that CsgG forms an ungated peptide diffusion channel. In PA$_{63}$, a model peptide diffusion channel, polypeptide passage depends on a ΔpH-driven Brownian ratchet that rectifies the diffusive steps in the translocation channel[20–22]. However, such proton gradients are not present at the outer membrane, requiring an alternative driving force. Whereas at elevated concentrations CsgG facilitates a non-selective diffusive leakage of periplasmic polypeptides, secretion is specific for CsgA under native conditions and requires the periplasmic factor CsgE[16,23]. In the presence of excess CsgE, purified CsgG forms a more slowly migrating species on native PAGE
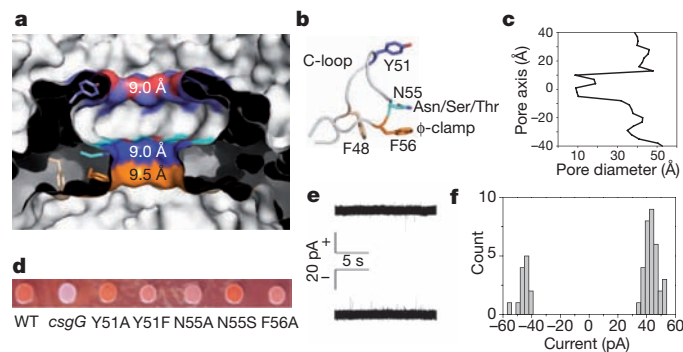
Figure 3 | CsgG channel constriction. a, Cross-section of CsgG channel constriction and its solvent-excluded diameters. **b**, The constriction is composed of three stacked concentric side-chain layers: Tyr 51, Asn 55 and Phe 56, preceded by Phe 48 from the periplasmic side. **c**, CsgG channel topology. **d**, Congo red binding of *E. coli* BW25141Δ*csgG* complemented with *csgG* (WT), empty vector or *csgG* carrying indicated constrictions mutants. Data are representative of six biological replicates. **e**, **f**, Representative single-channel current recordings (**e**) and conductance histogram (**f**) of CsgG reconstituted in planar phospholipid bilayers and measured under an electrical field of +50 mV ($n = 33$) or −50 mV ($n = 13$).
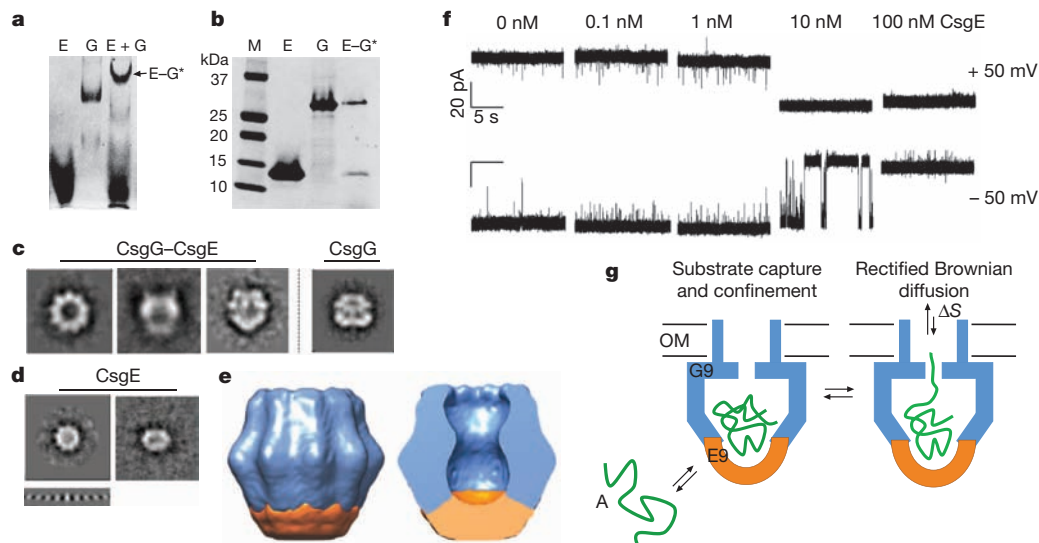
Figure 4 | Model of CsgG transport mechanism. a, Native PAGE of CsgE (E), CsgG (G) and CsgG supplemented with excess CsgE (E + G), showing the formation of a CsgG–CsgE complex (E–G*). Data are representative of seven experiments, encompassing four protein batches. **b**, SDS–PAGE of CsgE (E), CsgG (G) and the E–G* complex recovered from native PAGE. Data are representative of two repetitions. M, molecular mass markers. **c**, Selected class averages of CsgG–CsgE particles. From left to right: averaged top and side views visualized by cryo-EM, and comparison of negatively stained side views of CsgG-CsgE and CsgG nonamers. **d**, Cryo-EM averages of top and tilted side-viewed CsgE particles. Rotational autocorrelation shows nine-fold symmetry. **e**, Three-dimensional reconstruction of CsgG–CsgE (24 Å resolution, 1,221 single particles) shows a nonameric particle comprising CsgG (blue) and an additional density assigned as a CsgE nonamer (orange). **f**, Single-channel current recordings of PPB-reconstituted CsgG at +50 mV or −50 mV and supplemented with incremental concentrations of CsgE. Horizontal scale bars lie at 0 pA. **g**, Tentative model for CsgG-mediated protein secretion. CsgG and CsgE are proposed to form a secretion complex that entraps CsgA (discussed in Extended Data Fig. 9), generating an entropic potential over the channel. After capture of CsgA in the channel constriction, a ΔS-rectified Brownian diffusion facilitates the progressive translocation of the polypeptide across the outer membrane.

(Fig. 4a). SDS–PAGE analysis shows this new species consists of a CsgG–CsgE complex that is present in an equimolar stoichiometry (Fig. 4b). Cryo-electron microscopy (cryo-EM) visualization of CsgG–CsgE isolated by pull-down affinity purification revealed a nine-fold symmetrical particle corresponding to the CsgG nonamer and an additional capping density at the entrance to the periplasmic vestibule, similar in size and shape to a C_9 CsgE oligomer also observed by single-particle EM and size-exclusion chromatography (Fig. 4c–e and Extended Data Fig. 8). The location of the observed CsgG–CsgE contact interface was corroborated by blocking point mutations in CsgG helix 2 (Extended Data Fig. 8). In agreement with a capping function, single-channel recordings showed that CsgE binding to the translocator led to the specific silencing of its ion conductance (Fig. 4f and Extended Data Fig. 7). This CsgE capping of the channel seemed to be an all-or-none response in function of CsgE nomamer binding. At saturation, CsgE binding induced full blockage of the channel, independent of voltage sign, ruling out the possibility that purely electrophoretically or electroosmotically driven CsgE proteins clog the pore. At about 10 nM, an equilibrium between CsgE binding and dissociation events resulted in an intermittently blocked or fully open translocator. At 1 nM or below, transient (<1 ms) partial blockage events may have stemmed from short-lived encounters with monomeric CsgE.

Thus, CsgG and CsgE seem to form an encaging complex enclosing a central cavity of ~24,000 Å^3, reminiscent in appearance to the substrate-binding cavity and encapsulating lid structure seen in the GroEL chaperonin and GroES co-chaperonin[24]. The CsgG–CsgE enclosure would be compatible with the full or partial entrapment of the 129-residue CsgA. The caging of a translocation substrate has recently been observed in ABC toxins[25]. Spatial confinement of an unfolded polypeptide leads to a decrease in its conformational space, creating an entropic potential that has been shown to favour polypeptide folding in the case of chaperonins[24,26]. Similarly, we speculate that in curli transport the local high concentration and conformational confinement of curli subunits in the CsgG vestibule would generate an entropic free-energy gradient over the translocation channel (Fig. 4g). On capture into the constriction, the polypeptide chain

is then expected to move progressively outwards by Brownian diffusion, rectified by the entropic potential generated from the CsgE-mediated confinement and/or substrate trapping near the secretion channel. For full confinement in the pre-constriction cavity, the escape of an unfolded 129-residue polypeptide to the bulk solvent would correspond to an entropic free-energy release of up to ~80 kcal mol^{-1} (about 340 kJ mol^{-1}; ref. 27). The initial entropic cost of substrate docking and confinement are likely to be at least partly compensated for by binding energy released during assembly of the CsgG–CsgE–CsgA complex and an already lowered CsgA entropy in the periplasm. On theoretical grounds, three potential routes of CsgA recruitment to the secretion complex can be envisaged (Extended Data Fig. 9).

Curli-induced biofilms form a fitness and virulence factor in pathogenic Enterobacteriaceae[4,5]. Their unique secretion and assembly properties are also rapidly gaining interest for (bio)technological application[23,28,29]. Our structural characterization and biochemical study of two key secretion components provide a tentative model of an iterative mechanism for the membrane translocation of unfolded protein substrates in the absence of a hydrolysable energy source, a membrane potential or an ion gradient (Fig. 4e and Extended Data Fig. 9). The full validation and deconstruction of the contributing factors in the proposed secretion model will require the *in vitro* reconstitution of the translocator to allow transport kinetics to be followed accurately at the single-molecule level.

1. Olsen, A., Jonsson, A. & Normark, S. Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli. Nature* **338,** 652–655 (1989).
2. Collinson, S. K. *et al.* Thin, aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin. *J. Bacteriol.* **175,** 12–18 (1993).

3. Dueholm, M. S., Albertsen, M., Otzen, D. & Nielsen, P. H. Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. *PLoS ONE* **7,** e51274 (2012).
4. Cegelski, L. *et al.* Small-molecule inhibitors target *Escherichia coli* amyloid biogenesis and biofilm formation. *Nature Chem. Biol.* **5,** 913–919 (2009).
5. Herwald, H. *et al.* Activation of the contact-phase system on bacterial surfaces—a clue to serious complications in infectious diseases. *Nature Med.* **4,** 298–302 (1998).
6. Hammar, M., Arnqvist, A., Bian, Z., Olsen, A. & Normark, S. Expression of two *csg* operons is required for production of fibronectin- and Congo red-binding curli polymers in *Escherichia coli* K-12. *Mol. Microbiol.* **18,** 661–670 (1995).
7. Chapman, M. R. *et al.* Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295,** 851–855 (2002).
8. Wang, X., Smith, D. R., Jones, J. W. & Chapman, M. R. *In vitro* polymerization of a functional *Escherichia coli* amyloid protein. *J. Biol. Chem.* **282,** 3713–3719 (2007).
9. Dueholm, M. S. *et al.* Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation. *Biochemistry* **50,** 8281–8290 (2011).
10. Hung, C. *et al. Escherichia coli* biofilms have an organized and complex extracellular matrix structure. *MBio* **4,** e00645–13 (2013).
11. Hammar, M., Bian, Z. & Normark, S. Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli. Proc. Natl Acad. Sci. USA* **93,** 6562–6566 (1996).
12. Bian, Z. & Normark, S. Nucleator function of CsgB for the assembly of adhesive surface organelles in *Escherichia coli. EMBO J.* **16,** 5827–5836 (1997).
13. Loferer, H., Hammar, M. & Normark, S. Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. *Mol. Microbiol.* **26,** 11–23 (1997).
14. Robinson, L. S., Ashman, E. M., Hultgren, S. J. & Chapman, M. R. Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol. Microbiol.* **59,** 870–881 (2006).
15. Nenninger, A. A., Robinson, L. S. & Hultgren, S. J. Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc. Natl Acad. Sci. USA* **106,** 900–905 (2009).
16. Nenninger, A. A. *et al.* CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol. Microbiol.* **81,** 486–499 (2011).
17. Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* **65,** 239–259 (2011).
18. Iacovache, I., Bischofberger, M. & van der Goot, F. G. Structure and assembly of pore-forming proteins. *Curr. Opin. Struct. Biol.* **20,** 241–246 (2010).
19. Goyal, P., Van Gerven, N., Jonckheere, W. & Remaut, H. Crystallization and preliminary X-ray crystallographic analysis of the curli transporter CsgG. *Acta Crystallogr. F* **69,** 1349–1353 (2013).
20. Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309,** 777–781 (2005).
21. Janowiak, B. E., Fischer, A. & Collier, R. J. Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen. *J. Biol. Chem.* **285,** 8130–8137 (2010).
22. Feld, G. K., Brown, M. J. & Krantz, B. A. Ratcheting up protein translocation with anthrax toxin. *Protein Sci.* **21,** 606–624 (2012).
23. Van Gerven, N. *et al.* Secretion and functional display of fusion proteins through the curli biogenesis pathway. *Mol. Microbiol.* **91,** 1022–1035 (2014).
24. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107,** 223–233 (2001).
25. Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. & Lott, J. S. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501,** 547–550 (2013).
26. Takagi, F., Koga, N. & Takada, S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl Acad. Sci. USA* **100,** 11367–11372 (2003).
27. Zhou, H. X. Protein folding in confined and crowded environments. *Arch. Biochem. Biophys.* **469,** 76–82 (2008).
28. Chen, A. Y. *et al.* Synthesis and patterning of tunable multiscale materials with engineered cells. *Nature Mater.* **13,** 515–523 (2014).
29. Sivanathan, V. & Hochschild, A. A bacterial export system for generating extracellular amyloid aggregates. *Nature Protocols* **8,** 1381–1390 (2013).

METHODS

Cloning and strains. Expression constructs for the production of outer membrane localized C-terminally StrepII-tagged CsgG (pPG1) and periplasmic C-terminally StrepII-tagged CsgG$_{C1S}$ (pPG2) have been described in ref. 19. For selenomethionine labelling, StrepII-tagged CsgG$_{C1S}$ was expressed in the cytoplasm because of increased yields. Therefore, pPG2 was altered to remove the N-terminal signal peptide using inverse PCR with primers 5′-TCT TTA AC CGC CCC GCC TAA AG-3′ (forward) and 5′-CAT TTT TTG CCC TCG TTA TC-3′ (reverse) (pPG3). For phenotypic assays, a *csgG* deletion mutant of *E. coli* BW25141 (*E. coli* NVG2) was constructed by the method described in ref. 30 (with primers 5′-AAT AAC TCA ACC GAT TTT TAA GCC CCA GCT TCA TAA GGA AAA TAA TCG TGT AGG CTG GAG CTG CTT C-3′ and 5′-CGC TTA AAC AGT AAA ATG CCG GAT GAT AAT TCC GGC TTT TTT ATC TGC ATA TGA ATA TCC TCC TTA G-3′). The various CsgG substitution mutants used for Cys accessibility assays and for phenotypic probing of the channel constriction were constructed by site-directed mutagenesis (QuikChange protocol; Stratagene) starting from pMC2, a pTRC99a vector containing *csgG* under control of the *trc* promoter[14].

Protein expression and purification. CsgG and CsgG$_{C1S}$ were expressed and purified as described[19]. In brief, CsgG was recombinantly produced in *E. coli* BL 21 (DE3) transformed with pPG1 and extracted from isolated outer membranes with the use of 1% n-dodecyl-β-D-maltoside (DDM) in buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT)). StrepII-tagged CsgG was loaded onto a 5 ml Strep-Tactin Sepharose column (Iba GmbH) and detergent-exchanged by washing with 20 column volumes of buffer A supplemented with 0.5% tetraethylene glycol monooctyl ether (C8E4; Affymetrix) and 4 mM lauryldimethylamine-*N*-oxide (LDAO; Affymetrix). The protein was eluted by the addition of 2.5 mM D-desthiobiotin and concentrated to 5 mg ml^{-1} for crystallization experiments. For selenomethionine labelling, CsgG$_{C1S}$ was produced in the Met auxotrophic strain B834 (DE3) transformed with pPG3 and grown on M9 minimal medium supplemented with 40 mg l^{-1} L-selenomethionine. Cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM DTT, supplemented with cOmplete Protease Inhibitor Cocktail (Roche) and disrupted by passage through a TS series cell disruptor (Constant Systems Ltd) operated at 20×10^3 lb in^{-2}. Labelled CsgG$_{C1S}$ was purified as described[19]. DTT (5 mM) was added throughout the purification procedure to avoid oxidation of selenomethionine.

CsgE was produced in *E. coli* NEBC2566 cells harbouring pNH27 (ref. 16). Cell lysates in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 25 mM imidazole, 5% (v/v) glycerol were loaded on a HisTrap FF (GE Healthcare). CsgE-his was eluted with a linear gradient to 500 mM imidazole in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% (v/v) glycerol buffer. Fractions containing CsgE were supplemented with 250 mM (NH$_4$)$_2$SO$_4$ and applied to a 5 ml HiTrap Phenyl HP column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 250 mM (NH$_4$)$_2$SO$_4$, 5% (v/v) glycerol. A linear gradient to 20 mM Tris-HCl pH 8.0, 10 mM NaCl, 5% (v/v) glycerol was applied for elution. CsgE containing fractions were loaded onto a Superose 6 Prep Grade 10/600 (GE Healthcare) column equilibrated in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5% (v/v) glycerol.

In-solution oligomeric state assessment. About 0.5 mg each of detergent-solubilized CsgG (0.5% C8E4, 4 mM LDAO) and CsgG$_{C1S}$ were applied to a Superdex 200 10/300 GL analytical gel filtration column (GE Healthcare) equilibrated with 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT, 4 mM LDAO and 0.5% C8E4 (CsgG) or with 25 mM Tris-HCl pH 8.0, 200 mM NaCl (CsgG$_{C1S}$), and run at 0.7 ml min^{-1}. The column elution volumes were calibrated with bovine thyroglobulin, bovine γ-globulin, chicken ovalbumin, horse myoglobulin and vitamin B$_{12}$ (Bio-Rad) (Extended Data Fig. 2). Membrane-extracted CsgG, 20 μg of the detergent-solubilized protein was also run on 3–10% blue native PAGE using the procedure described in ref. 31 (Extended Data Fig. 2). NativeMark (Life Technologies) unstained protein standard (7 μl) was used for molecular mass estimation.

Crystallization, data collection and structure determination. Selenomethionine-labelled CsgG$_{C1S}$ was concentrated to 3.8 mg ml^{-1} and crystallized by sitting-drop vapour diffusion against a solution containing 100 mM sodium acetate pH 4.2, 8% PEG 4000 and 100 mM sodium malonate pH 7.0. Crystals were incubated in crystallization buffer supplemented with 15% glycerol and flash-frozen in liquid nitrogen. Detergent-solubilized CsgG was concentrated to 5 mg ml^{-1} and crystallized by hanging-drop vapour diffusion against a solution containing 100 mM Tris-HCl pH 8.0, 8% PEG 4000, 100 mM NaCl and 500 mM MgCl$_2$. Crystals were flash-frozen in liquid nitrogen and cryoprotected by the detergent present in the crystallization solution. For optimization of crystal conditions and screening for crystals with good diffraction quality, crystals were analysed on beamlines Proxima-1 and Proxima-2a (Soleil, France), PX-I (Swiss Light Source, Switzerland), I02, I03, I04 and I24 (Diamond Light Source, UK) and ID14eh2, ID23eh1 and ID23eh2 (ESRF, France). Final diffraction data used for structure determination of CsgG$_{C1S}$ and CsgG were collected at beamlines I04 and I03, respectively (see Extended Data Fig. 10a for data collection and refinement statistics). Diffraction data for CsgG$_{C1S}$ were processed

using Xia2 and the XDS package[32,33]. Crystals of CsgG$_{C1S}$ belonged to space group *P*1 with unit cell dimensions of a = 101.3 Å, b = 103.6 Å, c = 141.7 Å, α = 111.3°, β = 90.5°, γ = 118.2°, containing 16 protein copies in the asymmetric unit. For structure determination and refinement, data collected at 0.9795 Å wavelength were truncated at 2.8 Å on the basis of an $I/\sigma I$ cutoff of 2 in the highest-resolution shell. The structure was solved using experimental phases calculated from a single anomalous dispersion (SAD) experiment. A total of 92 selenium sites were located in the asymmetric unit by using ShelxC and ShelxD[34], and were refined and used for phase calculation with Sharp[35] (phasing power 0.79, figure of merit (FOM) 0.25). Experimental phases were density modified and averaged by non-crystallographic symmetry (NCS) using Parrot[36] (Extended Data Fig. 10; FOM 0.85). An initial model was built with Buccaneer[37] and refined by iterative rounds of maximum-likelihood refinement with Phenix refine[38] and manual inspection and model (re)building in Coot[39]. The final structure contained 28,853 atoms in 3,700 residues belonging to 16 CsgG$_{C1S}$ chains (Extended Data Fig. 2), with a molprobity[40] score of 1.34; 98% of the residues lay in favoured regions of the Ramachandran plot (99.7% in allowed regions). Electron density maps showed no unambiguous density corresponding to possible solvent molecules, and no water molecules or ions were therefore built in. Sixteenfold NCS averaging was maintained throughout refinement, using strict and local NCS restraints in early and late stages of refinement, respectively.

Diffraction data for CsgG were collected from a single crystal at 0.9763 Å wavelength and were indexed and scaled, using the XDS package[32,33], in space group C2 with unit-cell dimensions a = 161.7 Å, b = 372.3 Å, c = 161.8 Å and β = 92.9°, encompassing 18 CsgG copies in the asymmetric unit and a 72% solvent content. Diffraction data for structure determination and refinement were elliptically truncated to resolution limits of 3.6 Å, 3.7 Å and 3.8 Å along reciprocal cell directions a*, b* and c* and scaled anisotropically with the Diffraction Anisotropy Server[41]. Molecular replacement using the CsgG$_{C1S}$ monomer proved unsuccessful. Analysis of the self rotation function revealed D_9 symmetry in the asymmetric unit (not shown). On the basis of on the CsgG$_{C1S}$ structure, a nonameric search model was generated in the assumption that after going from a C_8 to C_9 oligomer, the interprotomer arc at the particle circumference would stay approximately the same as the interprotomer angle changed from 45° to 40°, giving a calculated increase in radius of about 4 Å. Using the calculated nonamer as search model, a molecular replacement solution containing two copies was found with Phaser[42]. Inspection of density-modified and NCS-averaged electron density maps (Parrot[36]; Extended Data Fig. 10) allowed manual building of the TM1 and TM2 and remodelling of adjacent residues in the protein core, as well as the building of residues 2–18, which were missing from the CsgG$_{C1S}$ model and linked the α1 helix to the N-terminal lipid anchor. Refinement of the CsgG model was performed with Buster-TNT[43] and Refmac5 (ref. 44) for initial and final refinement rounds, respectively. Eighteenfold local NCS restraints were applied throughout refinement, and Refmac5 was run employing a jelly-body refinement with sigma 0.01 and hydrogen-bond restraints generated by Prosmart[45]. The final structure contained 34,165 atoms in 4,451 residues belonging to 18 CsgG chains (Extended Data Fig. 2), with a molprobity score of 2.79; 93.0% of the residues lay in favoured regions of the Ramachandran plot (99.3% in allowed regions). No unambiguous electron density corresponding the N-terminal lipid anchor could be discerned.

Congo red assay. For analysis of Congo red binding, a bacterial overnight culture grown at 37 °C in Lysogeny Broth (LB) was diluted in LB medium until a D_{600} of 0.5 was reached. A 5 μl sample was spotted on LB agar plates supplemented with ampicillin (100 mg l^{-1}), Congo red (100 mg l^{-1}) and 0.1% (w/v) isopropyl β-D-thiogalactoside (IPTG). Plates were incubated at room temperature (20–22 °C) for 48 h to induce curli expression. The development of the colony morphology and dye binding were observed at 48 h.

Cysteine accessibility assays. Cysteine mutants were generated in pMC2 using site-directed mutagenesis and expressed in *E. coli* LSR12 (ref. 7). Bacterial cultures grown overnight were spotted onto LB agar plates containing 1 mM IPTG and 100 mg l^{-1} ampicillin. Plates were incubated at room temperature and cells were scraped after 48 h, resuspended in 1 ml of PBS and normalized using D_{600}. The cells were lysed by sonication and centrifuged for 20 s at 3,000g at 4 °C to remove unbroken cells from cell lysate and suspended membranes. Proteins in the supernatant were labelled with 15 mM methoxypolyethylene glycol-maleimide (MAL-PEG 5 kDa) for 1 h at room temperature. The reaction was stopped with 100 mM DTT and centrifuged at 40,000 r.p.m. (~100,000g) in a 50.4 Ti rotor for 20 min at 4 °C to pellet total membranes. The pellet was washed with 1% sodium lauroyl sarcosine to solubilize cytoplasmic membranes and centrifuged again. The resulting outer membranes were resuspended and solubilized using PBS containing 1% DDM. Metal-affinity pulldowns with nickel beads were used for SDS–PAGE and anti-His western blots. *E. coli* LSR12 cells with empty pMC2 vector were used as negative control.

ATR–FTIR spectroscopy. ATR–FTIR measurements were performed on an Equinox 55 infrared spectrophotometer (Bruker), continuously purged with dried air, equipped with a liquid-nitrogen-refrigerated mercury cadmium telluride detector

and a Golden Gate reflectance accessory (Specac). The internal reflection element was a diamond crystal (2 mm × 2 mm) and the beam incidence angle was 45°. Each purified protein sample (1 μl) was spread at the surface of the crystal and dried under a gaseous nitrogen flow to form a film. Each spectrum, recorded at 2 cm^{-1} resolution, was an average of 128 accumulations for improved signal-to-noise ratio. All the spectra were treated with water vapour contribution subtraction, smoothed at a final resolution of 4 cm^{-1} by apodization and normalized on the area of the Amide I band (1,700–1,600 cm^{-1}) to allow their comparison[46].

Negative stain EM and symmetry determination. Negative stain EM was used to monitor in-solution oligomerization states of CsgG, CsgG$_{C1S}$ and CsgE. CsgE, CsgG$_{C1S}$ and amphipol-bound CsgG were adjusted to a concentration of 0.05 mg ml^{-1} and applied to glow-discharged carbon-coated copper grids (CF-400; Electron Microscopy Sciences). After 1 min incubation, samples were blotted, then washed and stained in 2% uranyl acetate. Images were collected on a Tecnai T12 BioTWIN LaB6 microscope operating at a voltage of 120 kV, at a nominal magnification of ×49,000 and defocus between 800 and 2,000 nm. Contrast transfer function (CTF), phase flipping and particle selection were performed as described for cryo-EM. For membrane-extracted CsgG, octadecameric particles (1,780 in all) were analysed separately from nonamers and top views. For purified CsgE a total of 2,452 particles were analysed. In all cases, after normalization and centring, images were classified using IMAGIC-4D as described in the cryo-EM section. The best classes corresponding to characteristic views were selected for each set of particles. Symmetry determination of CsgG top views was performed using the best class averages with roughly 20 images per class. The rotational autocorrelation function was calculated using IMAGIC and plotted.

CsgG–CsgE complex formation. For CsgG–CsgE complex formation, the solubilizing detergents in purified CsgG were exchanged for Amphipols A8-35 (Anatrace) by adding 120 μl of CsgG (24 mg ml^{-1} protein in 0.5% C8E4, 4 mM LDAO, 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT) to 300 μl of detergent-destabilized liposomes (1 mg ml^{-1} 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) and 0.4% LDAO) and incubating for 5 min on ice before the addition of 90 μl of A8-35 amphipols at 100 mg ml^{-1} stock. After an additional 15 min incubation on ice, the sample was loaded on a Superose 6 10/300 GL (GE Healthcare) column and gel filtration was performed in 200 mM NaCl, 2.5% xylitol, 25 mM Tris-HCl pH 8, 0.2 mM DTT. An equal volume of purified monomeric CsgE in 200 mM NaCl, 2.5% xylitol, 25 mM Tris-HCl pH 8, 0.2 mM DTT was added to the amphipol-solubilized CsgG at final protein concentrations of 15 and 5 μM for CsgE and CsgG, respectively, and the sample was run at 125 V at 18 °C on a 4.5% native PAGE in 0.5 × TBE buffer. For the second, denaturing dimension, the band corresponding to the CsgG–CsgE complex was cut out of unstained lanes run in parallel on the same gel, boiled for 5 min in Laemmli buffer (60 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.01% bromophenol blue) and run on 4–20% SDS–PAGE. Purified CsgE and CsgG were run alongside the complex as control samples. Gels were stained with InstantBlue Coomassie for visual inspection or SYPRO orange for stoichiometry assessment of the CsgG–CsgE complex by fluorescence detection (Typhoon FLA 9000) of the CsgE and CsgG bands on SDS–PAGE, yielding a CsgG/CsgE ratio of 0.97.

CsgG–CsgE Cryo-EM. Cryo-electron microscopy was used to determine the in-solution structure of the C_9 CsgG–CsgE complex. CsgG–CsgE complex prepared as described above was bound and eluted with buffer supplemented with 100 mM imidazole from a TALON cobalt metal affinity resin to remove unbound CsgG, and on elution it was immediately applied to Quantifoil R2/2 carbon coated grids (Quantifoil Micro Tools GmbH) that had been glow-discharged at 20 mA for 30 s. Samples were plunge-frozen in liquid nitrogen using an automated system (Leica) and observed under a FEI F20 microscope operating at a voltage of 200 kV, a nominal magnification of ×50,000 under low-dose conditions and a defocus range of 1.4–3 μm. Image frames were recorded on a Falcon II detector. The pixel size at the specimen level was 1.9 Å per pixel. The CTF parameters were assessed using CTFFIND3 (ref. 47), and the phase flipping was done in SPIDER[48]. Particles were automatically selected from CTF-corrected micrographs using BOXER (EMAN2; ref. 49). Images with an astigmatism of more than 10% were discarded. A total of 4,881 particles were selected from 75 micrographs and windowed into 128-pixel × 128-pixel boxes. Images were normalized to the same mean and standard deviation and high-pass filtered at a low-resolution cut-off of ~200 Å. They were centred and then subjected to a first round of MSA. An initial reference set was obtained using reference free classification in IMAGIC-4D (Image Science Software). The best classes corresponding to characteristic side views of the C_9 cylindrical particles were used as references for the MRA. For CsgG–CsgE complex, the first three-dimensional model was calculated from the best 125 characteristic views (with good contrast and well-defined features) encompassing 1,221 particles of the complex with orientations determined by angular reconstitution (Image Science Software). The three-dimensional map was refined by iterative rounds of MRA, MSA and anchor set refinement. The resolution was estimated to be 24 Å by

Fourier shell correlation (FSC) according to the 0.5 criteria level (Extended Data Fig. 7). Visualization of the map and figures was performed in UCSF Chimera[50].

Bile salt toxicity assay. Outer-membrane permeability was investigated by decreased growth on agar plates containing bile salts. Tenfold serial dilutions of *E. coli* LSR12 (ref. 7) cells (5 μl) harbouring both pLR42 (ref. 16) and pMC2 (ref. 14) (or derived helix 2 mutants) were spotted on McConkey agar plates containing 100 μg l^{-1} ampicillin, 25 μg l^{-1} chloramphenicol, 1 mM IPTG with or without 0.2% (w/v) L-arabinose. After incubation overnight at 37 °C, colony growth was examined.

Single-channel current recordings. Single-channel current recordings were performed using parallel high-resolution electrical recording with the Orbit 16 device from Nanion. In brief, horizontal bilayers of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) were formed over microcavities (of subpicolitre volume) in a 16-channel multielectrode cavity array (MECA) chip (Ionera)[51]. Both the *cis* and *trans* cavities above and below the bilayer contained 1.0 M KCl, 25 mM Tris-HCl pH 8.0. To insert channels into the membrane, CsgG dissolved in 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT, 0.5% C8E4, 5 mM LDAO was added to the *cis* compartment to a final concentration of 90–300 nM. To test the interaction of the CsgG channel with CsgE, a solution of the latter protein dissolved in 25 mM Tris-HCl pH 8.0, 150 mM NaCl was added to the *cis* compartment to final concentrations of 0.1, 1, 10 and 100 nM. Transmembrane currents were recorded at a holding potential of +50 mV and −50 mV (with the *cis* side grounded) using a Tecella Triton 16-channel amplifier at a low-pass filtering frequency of 3 kHz and a sampling frequency of 10 kHz. Current traces were analysed using the Clampfit of the pClamp suite (Molecular Devices). Plots were generated using Origin 8.6 (Microcal)[52].

Measured currents were compared with those calculated based on the pore dimensions of the CsgG X-ray structure, modelled to be composed of three segments: the transmembrane section, the periplasmic vestibule, and the inner channel constriction connecting the two. The first two segments were modelled to be of conical shape; the constriction was represented as a cylinder. The corresponding resistances R_1, R_2 and R_3, respectively, were calculated as

$$R_1 = L_1/(\pi D_1 d_1 \kappa)$$

$$R_2 = L_2/(\pi D_2 d_2 \kappa)$$

$$R_3 = L_3/(\pi d_1 d_2 \kappa)$$

where L_1, L_2 and L_3 are the axial lengths of the segments, measuring 3.5, 4.0 and 2.0 nm, respectively, and D_1, d_1, D_2 and d_2 are the maximum and minimum diameters of segments 1 and 2, measuring 4.0, 0.8, 3.5 and 0.8 nm, respectively. The conductivity κ has the macroscopic bulk value of 10.6 S m^{-1} for the wider conical segments. The conductivity was half this value for the narrow central constriction, owing to the reduced ion mobility, in line with findings for the OmpF pore of similar dimensions[53]. The current was calculated by inserting R_1, R_2 and R_3 and voltage $V = 50$ mV into

$$I = V/(R_1 + R_2 + R_3)$$

Access resistance also included in the calculations.

30. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
31. Swamy, M., Siegers, G. M., Minguet, S., Wollscheid, B. & Schamel, W. W. A. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for the identification and analysis of multiprotein complexes. *Sci. STKE* **2006,** pl4, http://dx.doi.org/10.1126/stke.3452006pl4 (2006).
32. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
33. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
34. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66,** 479–485 (2010).
35. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
36. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66,** 470–478 (2010).
37. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
40. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35** (Suppl 2), W375–W383 (2007).
41. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).

42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
43. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68,** 368–380 (2012).
44. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
45. Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. D* **68,** 404–417 (2012).
46. Goormaghtigh, E. & Ruysschaert, J. M. Subtraction of atmospheric water contribution in Fourier transform infrared spectroscopy of biological membranes and proteins. *Spectrochim. Acta* **50A,** 2137–2144 (1994).
47. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
48. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3,** 1941–1974 (2008).
49. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
50. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
51. Del Rio Martinez, J. M., Zaitseva, E., Petersen, S., Baaken, G. & Behrends, J. C. Automated formation of lipid membrane microarrays for ionic single molecule sensing with protein nanopores. *Small* http://dx.doi.org/10.1002/smll.201402016 (13 August 2014).
52. Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nature Biotechnol.* **18,** 1091–1095 (2000).
53. Im, W. & Roux, B. Ions and counterions in a biological channel: a molecular dynamics simulation of OmpF porin from Escherichia coli in an explicit membrane with 1 M KCl aqueous salt solution. *J. Mol. Biol.* **319,** 1177–1179 (2002).
54. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).
55. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958 (2010).
56. Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23,** 1875–1882 (2007).
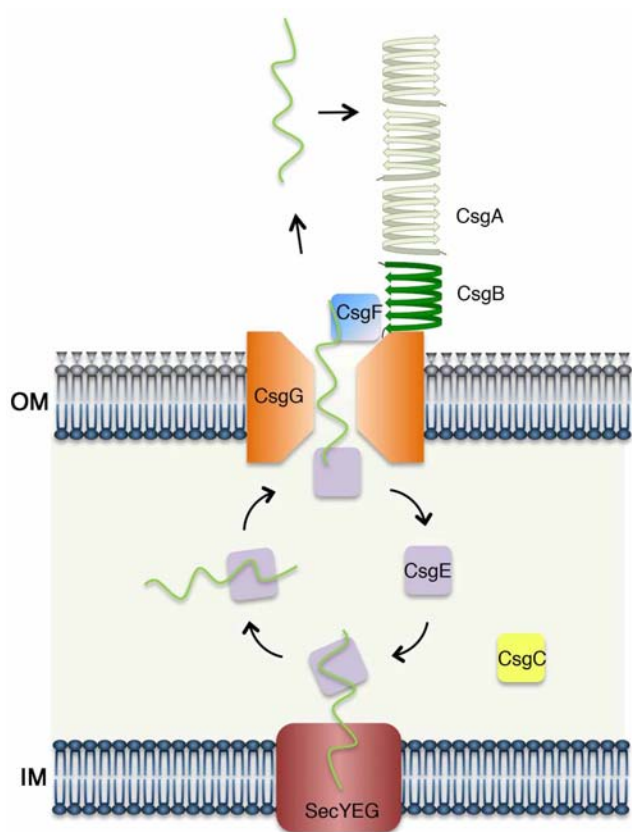
Extended Data Figure 1 | Curli biosynthesis pathway in *E. coli*. The major curli subunit CsgA (light green) is secreted from the cell as a soluble monomeric protein. The minor curli subunit CsgB (dark green) is associated with the outer membrane (OM) and acts as a nucleator for the conversion of CsgA from a soluble protein to amyloid deposit. CsgG (orange) assembles into an oligomeric curli-specific translocation channel in the outer membrane. CsgE (purple) and CsgF (light blue) form soluble accessory proteins required for productive CsgA and CsgB transport and deposition. CsgC forms a putative oxidoreductase of unknown function. All curli proteins have putative Sec signal sequences for transport across the cytoplasmic (inner) membrane (IM).
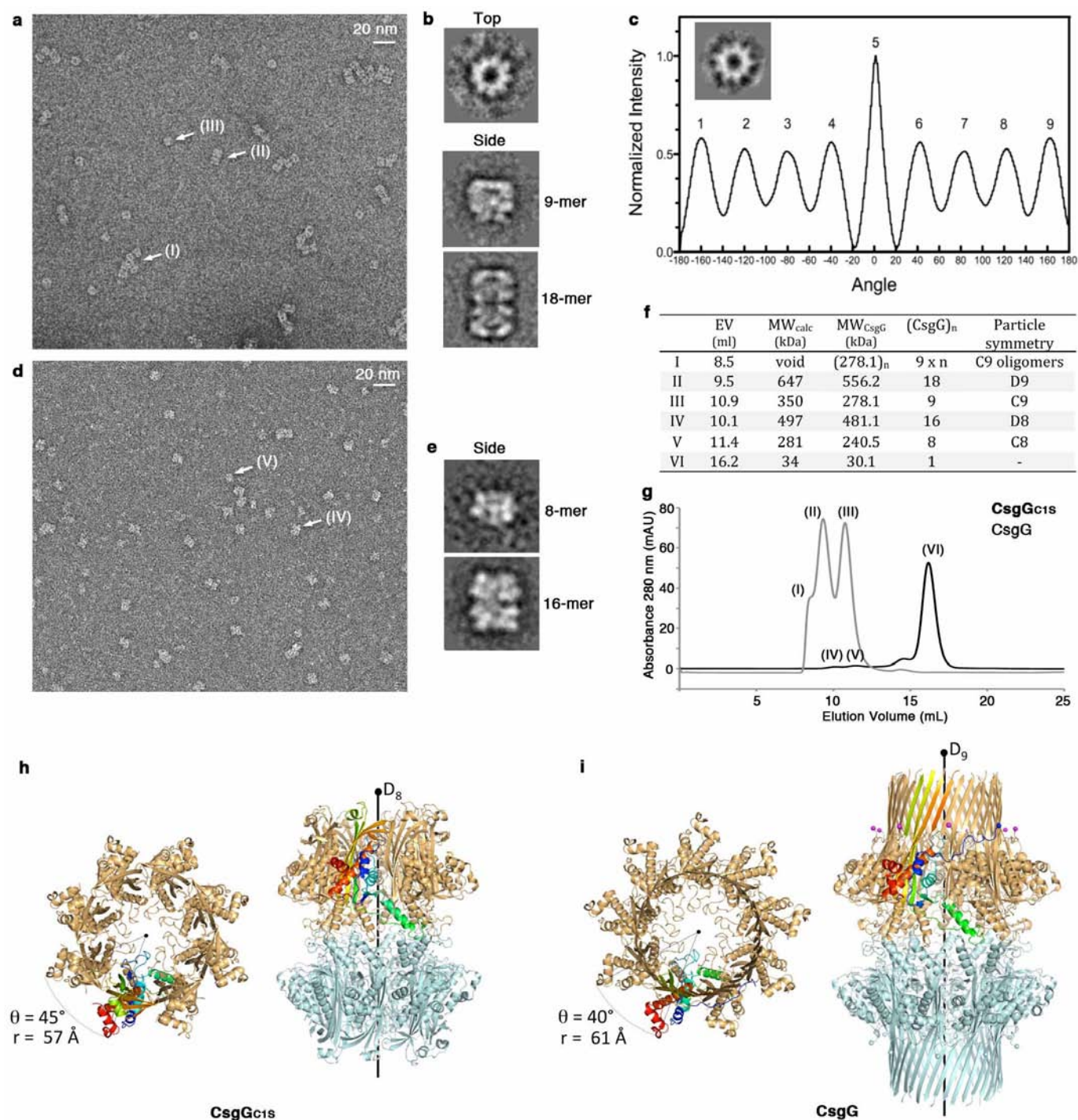
Extended Data Figure 2 | In-solution oligomerization states of CsgG and CsgG$_{C1S}$ analysed by size-exclusion chromatography and negative-stain electron microscopy. a, Raw negative-stain EM image of C8E4/LDAO-solubilized CsgG. Arrows indicate the different particle populations as labelled in the size exclusion profile shown in **g**, being (I) aggregates of CsgG nonamers, (II) CsgG octadecamers and (III) CsgG nonamers. Scale bar, 20 nm. **b**, Representative class average for top and side views of the indicated oligomeric states. **c**, Rotational autocorrelation function graph of LDAO-solubilized CsgG in top view, showing nine-fold symmetry. **d**, Raw negative-stain EM image of CsgG$_{C1S}$. Arrows indicate the hexadecameric (IV) and octameric (V) particles observed by size-exclusion chromatography in **g**. **e**, Representative class average for side views of CsgG$_{C1S}$ oligomers. No top views were observed for this construct. **f**, Table of elution volumes (EV) of CsgG$_{C1S}$ and CsgG particles observed by size-exclusion chromatography shown in **g**, calculated molecular mass (MW$_{calc}$), expected molecular mass (MW$_{CsgG}$) corresponding CsgG oligomerization state (CsgG$_n$) and the particles' symmetry as observed by negative-stain EM and X-ray crystallography. **g**, Size-exclusion chromatogram of CsgG$_{C1S}$ (black) and C8E4/LDAO-solubilized CsgG (grey) run on Superdex 200 10/300 GL (GE Healthcare). **h, i**, Ribbon representation of crystallized oligomers in top and side view, showing the D_8 hexadecamers for CsgG$_{C1S}$ (**h**) and D_9 octadecamers for membrane-extracted CsgG (**i**). One protomer is coloured in rainbow from N terminus (blue) to C terminus (red). The two C_8 octamers (CsgG$_{C1S}$) or C_9 nonamers (CsgG) that form the tail-to-tail dimers captured in the crystals are coloured blue and tan. r and θ give radius and interprotomer rotation, respectively.
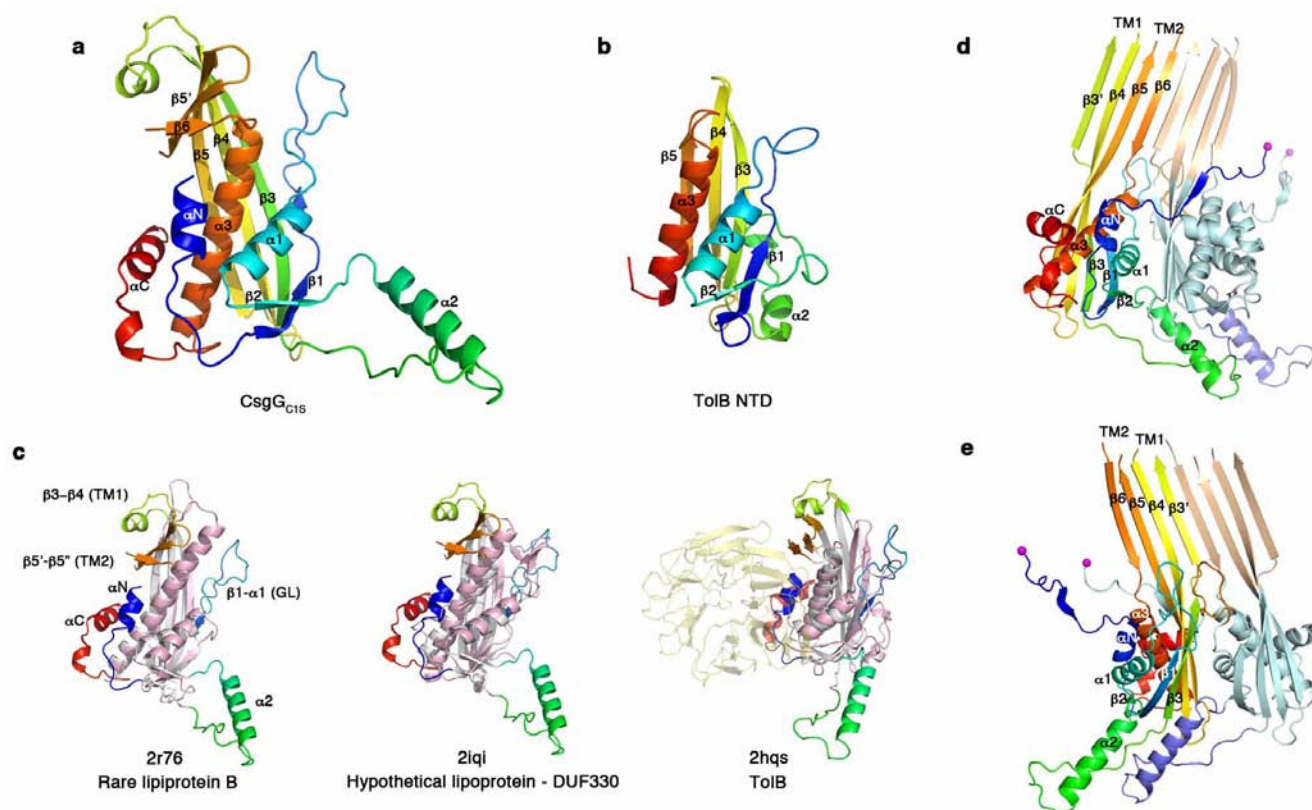
Extended Data Figure 3 | Comparison of CsgG with structural homologues and interprotomer contacts in CsgG. a, b, Ribbon diagram for the CsgG$_{C1S}$ monomer (for example CsgG in pre-pore conformation) (**a**) and the nucleotide-binding-domain-like domain of TolB (**b**) (PDB 2hqs), both coloured in rainbow from N terminus (blue) to C terminus (red). Common secondary structure elements are labelled equivalently. **c,** CsgG$_{C1S}$ (grey) in superimposition with, from left to right, *Xanthomonas campestris* rare lipoprotein B (PDB 2r76, coloured pink), *Shewanella oneidensis* hypothetical lipoprotein DUF330 (PDB 2iqi, coloured pink) and *Escherichia coli* TolB (PDB 2hqs, coloured pink and yellow for the N-terminal and β-propeller domains, respectively). CsgG-specific structural elements are labelled and coloured as in the upper left panel. **d, e,** Ribbon diagram of two adjacent protomers as found in the CsgG structure, viewed along the plane of the bilayer, either from outside (**c**) or inside (**d**) the oligomer. One protomer is shown in rainbow (dark blue to red) from N terminus to C terminus; a second protomer is shown in light blue (core domain), blue (helix 2) and tan (TM domain). Four main oligomerization interfaces are apparent: β6–β3′ main-chain interactions inside the β-barrel, the constriction loop (CL), side-chain packing of helix 1 (α1) against β1–β3–β4–β5, and helix–helix packing of helix 2 (α2). The 18-residue N-terminal loop connecting the lipid anchor (a magenta sphere shows the Cα position of Leu 2) and N-terminal helix (αN) is also seen to wrap over the adjacent two protomers. The projected position of the lipid anchor is expected to lie against the TM1 and TM2 hairpins of the +2 protomer (not shown for clarity).
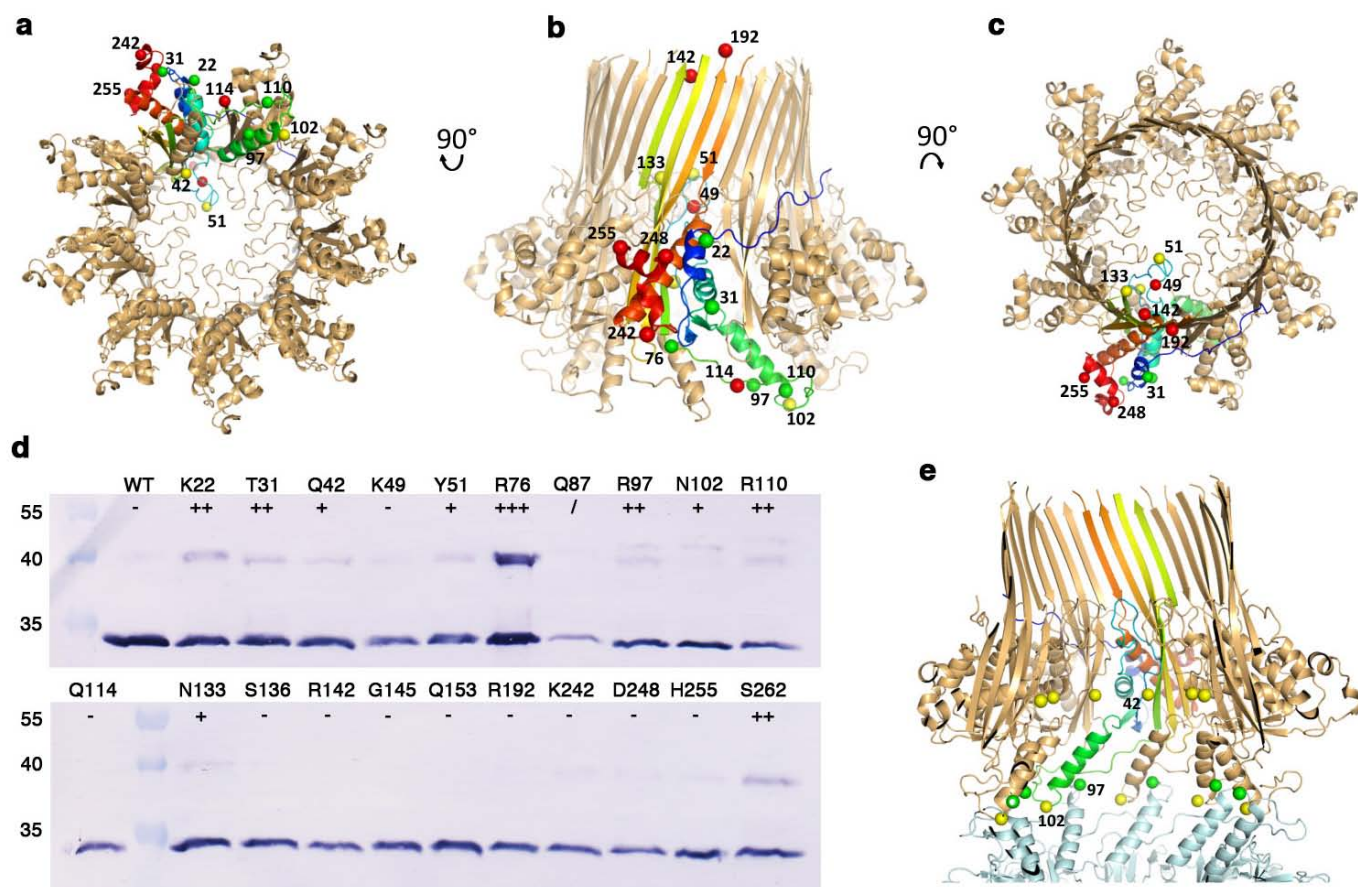
Extended Data Figure 4 | Cys accessibility assays for selected surface residues in the CsgG oligomers. a–c, Ribbon representation of CsgG nonamers shown in periplasmic (**a**), side (**b**) and extracellular (**c**) views. One protomer is coloured in rainbow from N terminus (blue) to C terminus (red). Cysteine substitutions are labelled and the equivalent locations of the S atoms are shown as spheres, coloured according to accessibility to MAL-PEG (5,000 Da) labelling in *E. coli* outer membranes. **d**, Western blot of MAL-PEG reacted samples analysed on SDS–PAGE, showing 5 kDa increase on MAL-PEG binding of the introduced cysteine. Accessible (++ and +++), moderately accessible (+) and inaccessible (−) sites are coloured green, orange and red, respectively, in **a–e**. For Arg 97 and Arg 110 a second species at 44 kDa is present, corresponding to a fraction of protein in which both the introduced

and native cysteine became labelled. Data are representative of four independent experiments from biological replicates. **e**, Side view of the dimerization interface in the D_9 octadecamer as present in the X-ray structure. Introduced cysteines in the dimerization interface or inside the lumen of the D_9 particle are labelled. In membrane-bound CsgG, these residues are accessible to MAL-PEG, demonstrating that the D_9 particles are an artefact of concentrated solutions of membrane-extracted CsgG and that the C_9 complex forms the physiologically relevant species. Residues in the C-terminal helix (αC; Lys 242, Asp 248 and His 255) are found to be inaccessible to poorly accessible, indicating that αC may form additional contacts with the *E. coli* cell envelope, possibly the peptidoglycan layer.
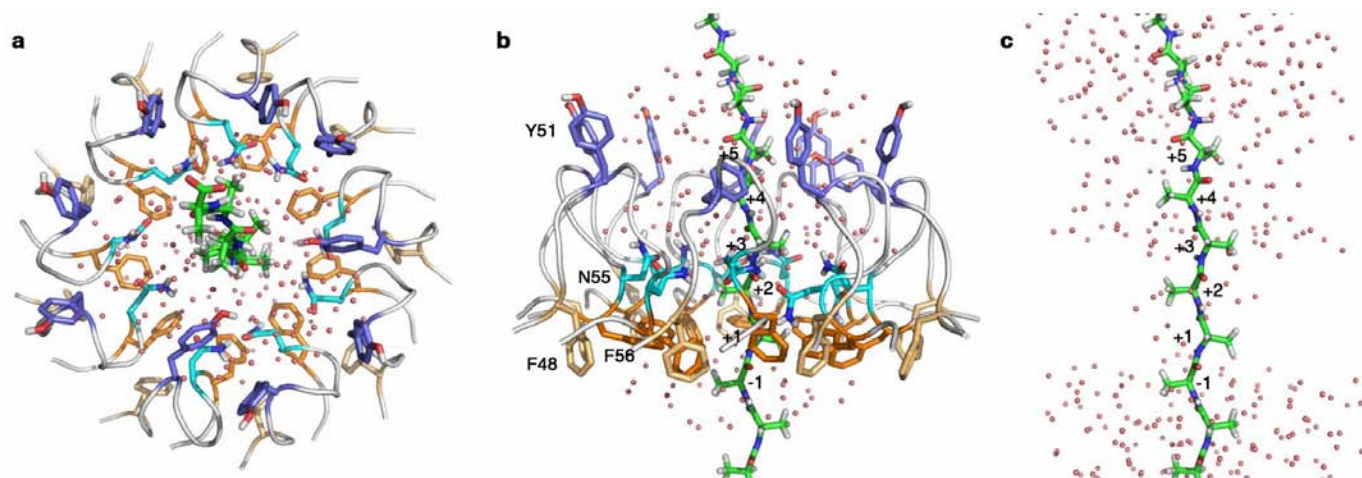
Extended Data Figure 5 | Molecular dynamics simulation of CsgG constriction with model polyalanine chain. a, b, Top (**a**) and side (**b**) views of the CsgG constriction modelled with a polyalanine chain threaded through the channel in an extended conformation, here shown in a C-terminal to N-terminal direction. Substrate passage through the CsgG transporter is itself not sequence specific[16,23]. For clarity, a polyalanine chain was used for modelling the putative interactions of a passing polypeptide chain. The modelled area is composed of nine concentric CsgG C-loops, each comprising residues 47–58. Side chains lining the constriction are shown in stick representation, with Phe 51 coloured slate blue, Asn 55 (amide-clamp) cyan, and Phe 48 and Phe 56 (ϕ-clamp) in light and dark orange, respectively. N, O and H atoms (only hydroxyl or side-chain amide H atoms are shown) are coloured blue, red and white, respectively. The polyalanine chain is coloured green, blue, red and white for C, N, O and H atoms, respectively.

Solvent molecules (water) within 10 Å of the polyalanine residues inside the constriction (residues labelled +1 to +5) are shown as red dots. **c,** Modelled solvation of the polyalanine chain, position as in **b** and with C-loops removed for clarity (shown solvent molecules are those within 10 Å of the full polyalanine chain). At the height of the amide-clamp and ϕ-clamp, the solvation of the polyalanine chain is reduced to a single water shell that bridges the peptide backbone and amide-clamp side chains. Most side chains in the Tyr 51 ring have rotated towards the solvent in comparison with their inward, centre-pointing position observed in the CsgG (and the CsgG$_{C1S}$) X-ray structure. The model is the result of a 40 ns all-atom explicit solvent molecular dynamics simulation with GROMACS[54] using the AMBER99SB-ILDN[55] force field and with the Cα atoms of the residues at the extremity of the C-loop (Gln 47 and Thr 58) positionally restricted.
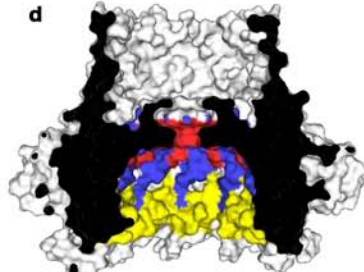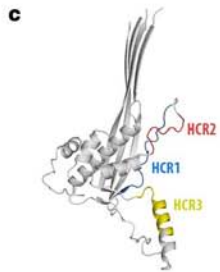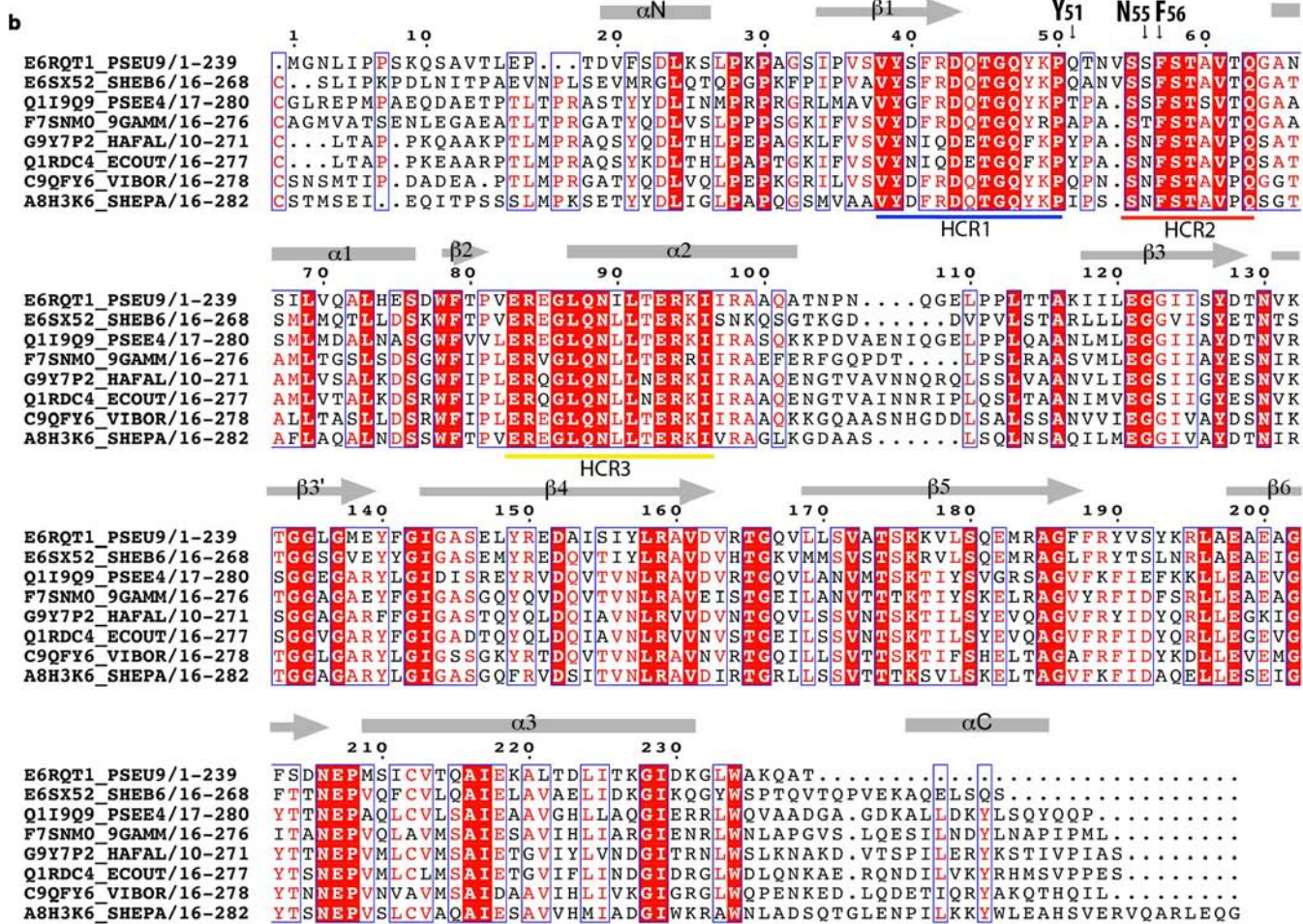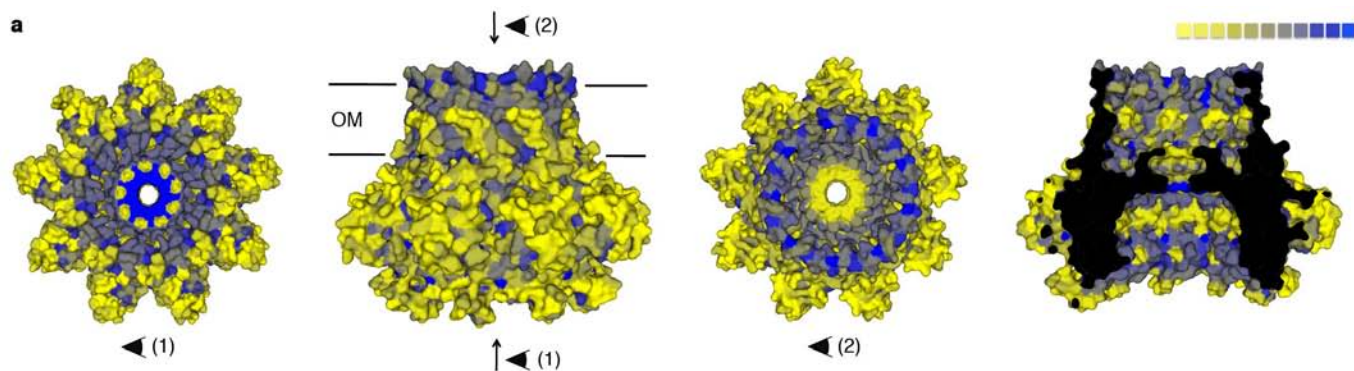
Extended Data Figure 6 | Sequence conservation in CsgG homologues.
a, Surface representation of the CsgG nonamer coloured according to sequence similarity (coloured yellow to blue from low to high conservation score)[56] and viewed from the periplasm (far left), the side (middle left), the extracellular milieu (middle right) or as a cross-sectional side view (far right). The figures show that the regions of highest sequence conservation map to the entry of the periplasmic vestibule, the vestibular side of the constriction loop and the luminal surface of the TM domain. **b**, Multiple sequence alignment of CsgG-like lipoproteins. The selected sequences were chosen from monophyletic clades across the phylogenetic three of CsgG-like sequences (not shown), to give a representative view of sequence diversity. Secondary structure elements are shown as arrows or bars for β-strands and α-helices, respectively, and are based on the *E. coli* CsgG crystal structure. **c, d**, CsgG protomer in secondary structure representation (**c**) and a cross-sectional side view (**d**) of the CsgG nonamer in surface representation, both coloured grey and with three continuous blocks of high sequence conservation coloured red (HCR1), blue (HCR2) and yellow (HCR3). HCR1 and HCR2 shape the vestibular side of the constriction loop; HCR3 corresponds to helix 2, lying at the entry of the periplasmic vestibule. Inside the constriction, Phe 56 is 100% conserved, whereas Asn 55 can be conservatively replaced by Ser or Thr, for example by a small polar side chain that can act as hydrogen-bond donor/acceptor. The concentric side-chain ring at the exit of the constriction (Tyr 51) is not conserved. The presence of the Phe-ring at the entrance of the constriction is topologically similar to the Phe 427-ring (referred to as the ϕ-clamp) in the anthrax protective antigen PA_{63}, in which it was shown to catalyse polypeptide capture and passage[20]. MST of *toxB* superfamily proteins reveals a conserved motif D(D/Q)(**F**)(S/N)S at the height of the Phe-ring. This is similar to the S(Q/N/T)(**F**)ST motif seen in curli-like transporters. Although an atomic resolution structure of PA_{63} in pore conformation is not yet available, available structures suggest the Phe-ring may similarly be followed by a conserved hydrogen-bond donor/acceptor (Ser/Asn 428) as a subsequent concentric ring in the translocation channel (note that the orientation of the element is inverted in both transporters).

a CsgG

b CsgG + CsgE

c CsgG D_9 double pore

d CsgD D_9 double pore + CsgE

10 nM CsgE +50 mV

100 nM CsgE +50 mV

−50mV

−50mV
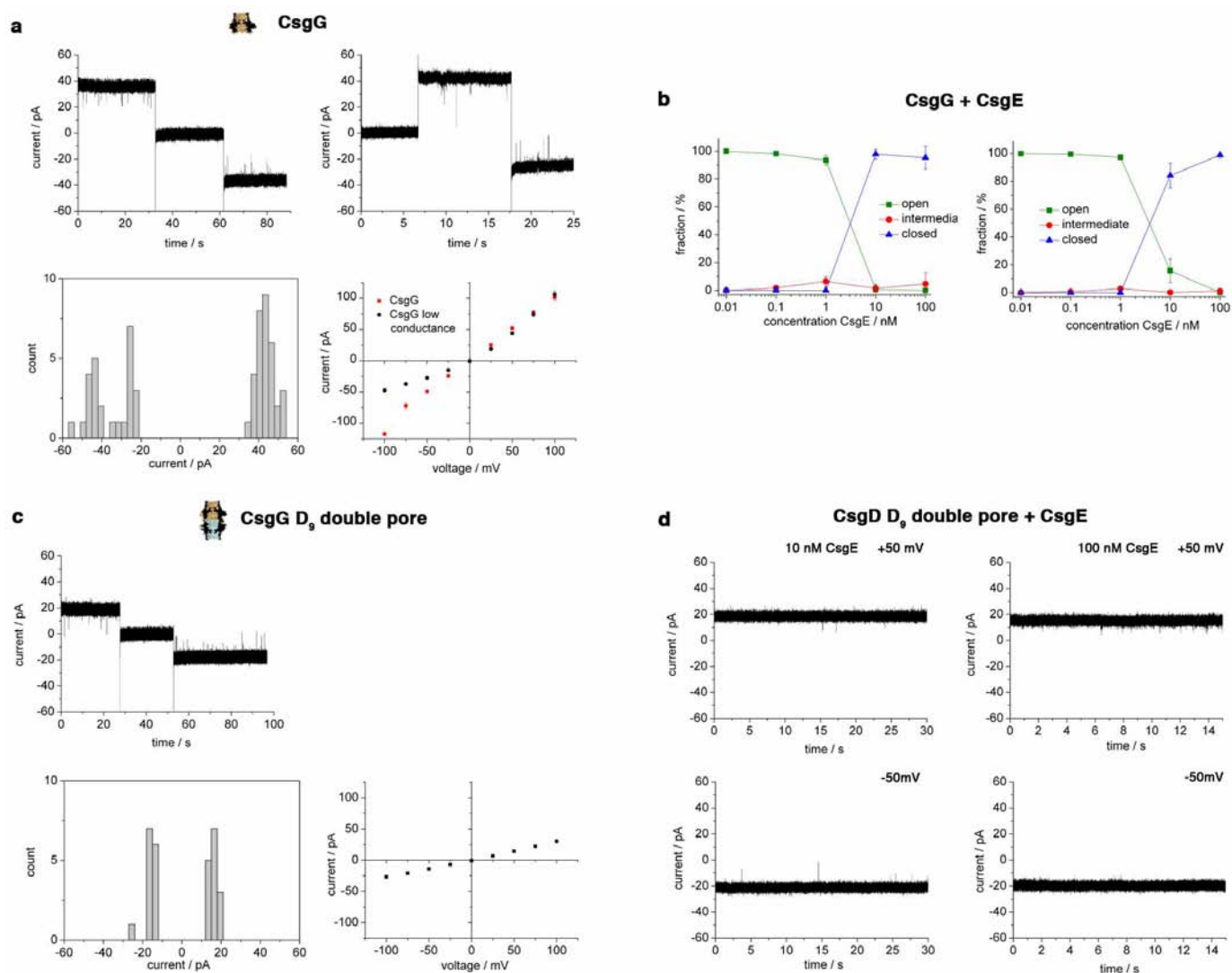
Extended Data Figure 7 | Single-channel current analysis of CsgG and CsgG + CsgE pores. a, Under negative field potential, CsgG pores show two conductance states. The upper left and right panels show a representative single-channel current trace of, respectively, the normal (measured at +50, 0 and −50 mV) and the low-conductance forms (measured at 0, +50 and −50 mV). No conversions between both states were observed during the total observation time ($n = 22$), indicating that the conductance states have long lifetimes (second to minute timescale). The lower left panel shows a current histogram for the normal and low-conductance forms of CsgG pores acquired at +50 and −50 mV ($n = 33$). I–V curves for CsgG pores with regular and low conductance are shown in the lower right panel. These data represent averages and standard deviations from at least four independent recordings. The nature or physiological existence of the low-conductance form is unknown. **b**, Electrophysiology of CsgG channels titrated with the periplasmic factor CsgE. The plots display the normalized occurrence, that is, the fractions of open, closed and intermediate-state channels, as a function of CsgE concentration. Open and closed states of CsgG are illustrated in Fig. 4f. Increasing the concentration of CsgE to more than 10 nM leads to the closure of CsgG pores. The effect occurs at +50 mV (left) and −50 mV (right), ruling out the possibility that the pore blockade is caused by electroosmosis or electrophoresis of CsgE (calculated p*I* 4.7) into the CsgG pore. An infrequent (<5%) intermediate state has roughly half the conductance of the open channel. It may represent CsgE-induced incomplete closures of the CsgG channel. Alternatively, it could represent the temporary formation of a CsgG dimer caused by the binding of residual CsgG monomer from the electrolyte solution to the membrane-embedded pore. The fractions for the three states were obtained from all-point histogram analysis of single-channel current

traces. The histograms yielded peak areas for up to three states, and the fraction for a given state was obtained by dividing the corresponding peak area by the sum of all other states in the recording. Under negative field potential, two open conductance states are discerned, similar to the observations for CsgG (see **a**). Because both open channel variations were blocked by higher CsgE concentrations, the 'open' traces in **b** combine both conductance forms. The data in the plot represent averages and standard deviations from three independent recordings. **c**, The crystal structure, size-exclusion chromatography and EM show that detergent extracted CsgG pores form non-native tail-to-tail stacked dimers (for example, two nonamers as D_9 particle; Extended Data Fig. 2) at higher protein concentration. These dimers can also be observed in single-channel recordings. The upper panel shows the single-channel current trace of stacked CsgG pores at +50, 0 and −50 mV (left to right). The lower left panel shows a current histogram of dimeric CsgG pores recorded at +50 and −50 mV. The experimental conductances of +16.2 ± 1.8 and –16.0 ± 3.0 pA ($n = 15$) at +50 and −50 mV, respectively, are near the theoretically calculated value of 23 pA. The lower right panel shows an I–V curve for the stacked CsgG pores. The data represent averages and standard deviations from six independent recordings. **d**, The ability of CsgE to bind and block stacked CsgG pores was tested by electrophysiology. Shown are single-channel current traces of stacked CsgG pore in the presence of 10 or 100 nM CsgE at +50 mV (upper) and −50 mV (lower). The current traces indicate that otherwise saturating concentrations of CsgE do not lead to pore closure for stacked CsgG dimers. These observations are in good agreement with the mapping of the CsgG–CsgE contact zone to helix 2 and the mouth of the CsgG periplasmic cavity as discerned by EM and site-directed mutagenesis (Fig. 4 and Extended Data Fig. 7).
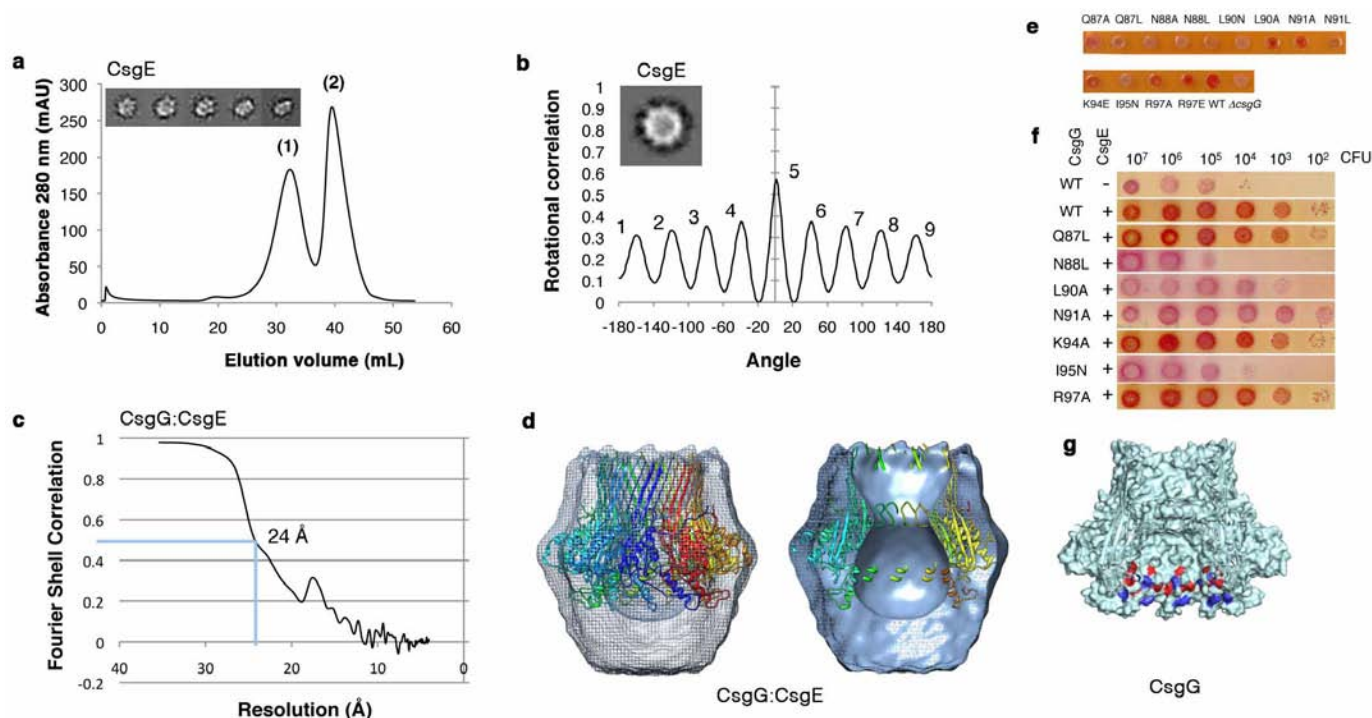
Extended Data Figure 8 | CsgE oligomer and CsgG–CsgE complex. a, Size-exclusion chromatography of CsgE (Superose 6, 16/600; running buffer 20 mM Tris-HCl pH 8, 100 mM NaCl, 2.5% glycerol) shows an equilibrium of two oligomeric states, 1 and 2, with an apparent molecular mass ratio of 9.16:1. Negative-stain EM inspection of peak 1 shows discrete CsgE particles (five representative class averages are shown in the inset, ordered by increasing tilt angles) compatible in size with nine CsgE copies. **b**, Selected class average of CsgE oligomer observed in top view by cryo-EM and its rotational autocorrelation show the presence of C_9 symmetry. **c**, FSC analysis of CsgG–CsgE cryo-EM model. Three-dimensional reconstruction achieved a resolution of 24 Å as determined by FSC at a threshold of 0.5 correlation using 125 classes corresponding to 1,221 particles. **d**, Overlay of CsgG–CsgE cryo-EM density and the CsgG nonamer observed in the X-ray structure. The overlays are shown viewed from the side as semi-transparent density (left) or as a cross-sectional view. **e**, Congo red binding of *E. coli* BW25141D*csgG* complemented with wild-type *csgG* (WT), empty vector (D*csgG*) or *csgG* helix 2 mutants (single amino acid replacements labelled in single-letter code). Data are representative of four biological replicates. **f**, Effect of bile salt toxicity on *E. coli* LSR12 complemented with *csgG* (WT) or on *csgG* carrying different helix 2 mutations, complemented with (+) or without (−) *csgE*. Tenfold serial dilution starting from 10^7 bacteria were spotted on McConkey agar plates. Expression of the CsgG pore in the outer membrane leads to an increased bile salt sensitivity that can be blocked by co-expression of CsgE (*n* = 6, three biological replicates, with two repetitions each). **g**, Cross-sectional view of CsgG X-ray structure in molecular surface representation. CsgG mutants without an effect on Congo red binding or toxicity are shown in blue; mutants that interfere with CsgE-mediated rescue of bile salt sensitivity are indicated in red.
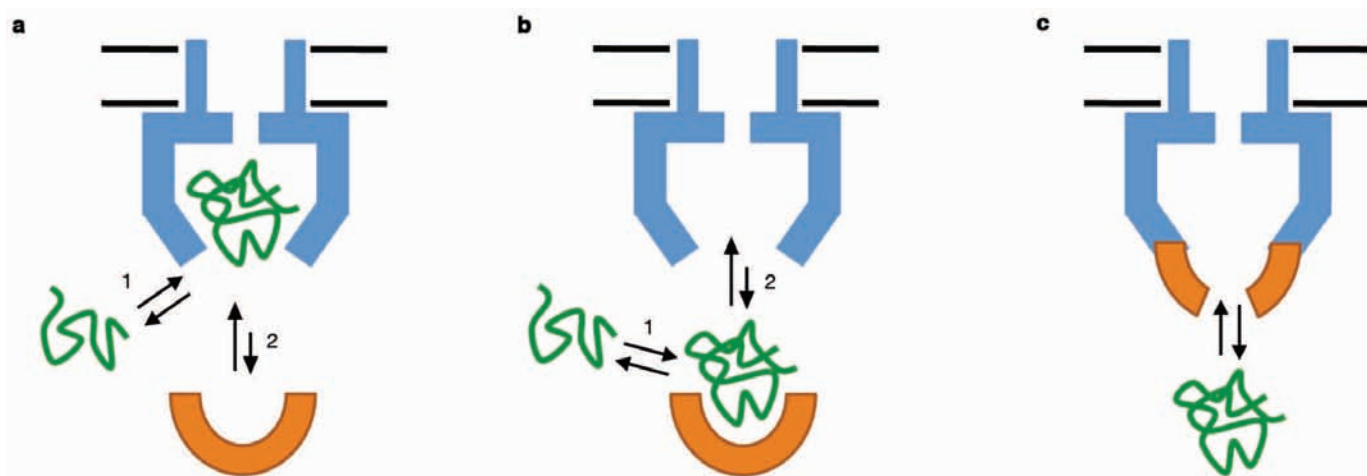
Extended Data Figure 9 | Assembly and substrate recruitment of the CsgG secretion complex. The curli transporter CsgG and the soluble secretion cofactor CsgE form a secretion complex with 9:9 stoichiometry that encloses a ~24,000 Å^3 chamber that is proposed to entrap the CsgA substrate and facilitate its entropy-driven diffusion across the outer membrane (OM; see the text and Fig. 4). On theoretical grounds, three putative pathways (**a–c**) for substrate recruitment and assembly of the secretion complex can be envisaged. **a**, A 'catch-and-cap' mechanism entails the binding of CsgA to the apo CsgG translocation channel (1), leading to a conformational change in the latter that exposes a high-affinity binding platform for CsgE binding (2). CsgE binding leads to capping of the substrate cage. On secretion of CsgA, CsgG would fall back into its low-affinity conformation, leading to CsgE dissociation and liberation of the secretion channel for a new secretion cycle. **b**, In a 'dock-and-trap' mechanism, periplasmic CsgA is first captured by CsgE (1), causing the latter to adopt a high-affinity complex that docks onto the CsgG translocation pore (2), enclosing CsgA in the secretion complex. CsgA binding could be directly to CsgE oligomers or to CsgE monomers, the latter leading to subsequent oligomerization and CsgG binding. Secretion of CsgA leads CsgE to fall back into its low-affinity conformation and to dissociate from the secretion channel. **c**, CsgG and CsgE form a constitutive complex, in which CsgE conformational dynamics cycle between open and closed forms in the course of CsgA recruitment and secretion. Currently published or available data do not allow us to discriminate between these the putative recruitment modes or derivatives thereof, or to put forward one of them.

a

Data collection and refinement statistics

	CsgG$_{C1S}$	CsgG
Data collection		
Space group	P1	C2
Cell dimensions		
a, b, c (Å)	101.3, 103.6, 141.7	161.9, 372.8, 161.9
α, β, γ (°)	111.3, 90.5, 118.2	90.0, 92.9, 90.0
Resolution (Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6)
		30 − 3.6 (a*), −3.7 (b*), −3.8 (c*) †
R_{meas} *	15.1 (81.8)	16.2 (90.6) †
$I/\sigma I$ *	9.82 (2.03)	6.80 (1.89) †
Completeness (%)*	98.7 (98.3)	91.57 (27.26)
		99.9 (99.1) †
Redundancy*	11.2 (7.0)	4.4 (4.3)
Wilson B (Å^2)	46.7	101.0
Refinement		
Resolution (Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6)
		30 − 3.6 (a*), −3.7 (b*), −3.8 (c*) †
No. reflections*	112419 (11159)	102130 (11094)
R_{work}/R_{free}	0.1881 / 0.2337	0.3024 / 0.3542
No. atoms		
Protein	28853	34165
Ligand/ion	0	0
Water	0	0
B-factors (Å^2)		
Protein	57.3	116.7
Ligand/ion		
Water		
R.m.s deviations		
Bond lengths (Å)	0.01	0.03
Bond angles (°)	1.31	1.87

Data statistics for CsgG$_{C1S}$ and membrane-extracted CsgG, collected from a single crystal each.
*Highest resolution shell is shown in parenthesis.
†Values corrected for anisotropic truncation along reciprocal directions a*, b* and c*.

b

c
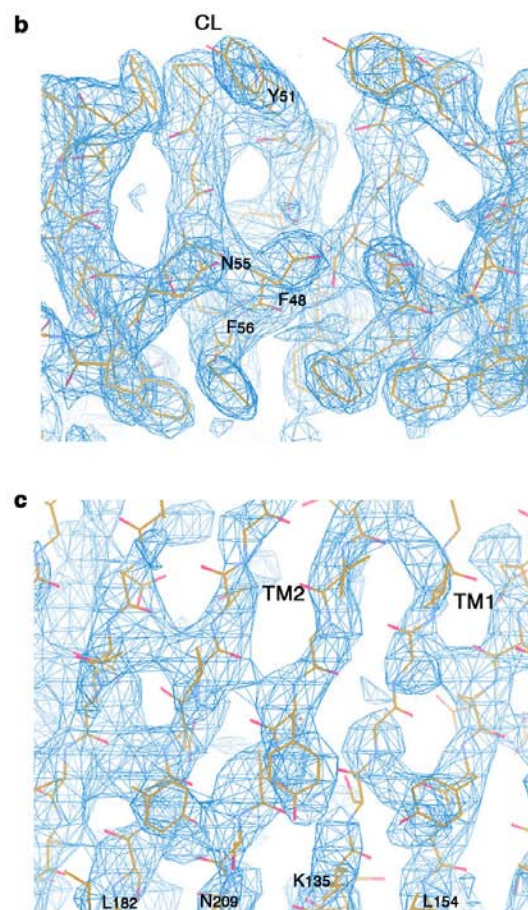
Extended Data Figure 10 | Data collection statistics and electron density maps of CsgG$_{C1S}$ and CsgG. a, Data collection statistics for CsgG$_{C1S}$ and CsgG X-ray structures. **b,** Electron density map at 2.8 Å for CsgG$_{C1S}$ calculated using NCS-averaged and density-modified experimental SAD phases, and contoured at 1.5σ. The map shows the region of the channel construction (CL; a single protomer is labelled) and is overlaid on the final refined model. **c,** Electron density map (resolutions 3.6, 3.7 and 3.8 Å along reciprocal vectors a*, b* and c*, respectively) in the CsgG TM domain region, calculated from NCS-averaged and density-modified molecular replacement phases (TM loops were absent from the input model); B-factor sharpened by −20 Å^2 and contoured at 1.0σ. The figure shows the TM1 (Lys 135–Leu 154) and TM2 (Leu 182–Asn 209) region of a single CsgG protomer, overlaid on the final refined model.

LETTER

Loss of signalling via Gα13 in germinal centre B-cell–derived lymphoma

Jagan R. Muppidi[1,2,3], Roland Schmitz[4], Jesse A. Green[1,3]†, Wenming Xiao[4], Adrien B. Larsen[5], Sterling E. Braun[1,3], Jinping An[1,3], Ying Xu[1,3], Andreas Rosenwald[6], German Ott[7,8], Randy D. Gascoyne[9], Lisa M. Rimsza[10], Elias Campo[11], Elaine S. Jaffe[12], Jan Delabie[13], Erlend B. Smeland[14,15], Rita M. Braziel[16], Raymond R. Tubbs[17], J. R. Cook[17], Dennis D. Weisenburger[18], Wing C. Chan[19,20], Nagarajan Vaidehi[5], Louis M. Staudt[4] & Jason G. Cyster[1,3]

Germinal centre B-cell-like diffuse large B-cell lymphoma (GCB-DLBCL) is a common malignancy, yet the signalling pathways that are deregulated and the factors leading to its systemic dissemination are poorly defined[1,2]. Work in mice showed that sphingosine-1-phosphate receptor-2 (S1PR2), a Gα12 and Gα13 coupled receptor, promotes growth regulation and local confinement of germinal centre B cells[3,4]. Recent deep sequencing studies of GCB-DLBCL have revealed mutations in many genes in this cancer, including in *GNA13* (encoding Gα13) and *S1PR2* (refs 5–7). Here we show, using *in vitro* and *in vivo* assays, that GCB-DLBCL-associated mutations occurring in S1PR2 frequently disrupt the receptor's Akt and migration inhibitory functions. Gα13-deficient mouse germinal centre B cells and human GCB-DLBCL cells were unable to suppress pAkt and migration in response to S1P, and Gα13-deficient mice developed germinal centre B-cell-derived lymphoma. Germinal centre B cells, unlike most lymphocytes, are tightly confined in lymphoid organs and do not recirculate. Remarkably, deficiency in Gα13, but not S1PR2, led to germinal centre B-cell dissemination into lymph and blood. GCB-DLBCL cell lines frequently carried mutations in the Gα13 effector *ARHGEF1*, and Arhgef1 deficiency also led to germinal centre B-cell dissemination. The incomplete phenocopy of Gα13- and S1PR2 deficiency led us to discover that P2RY8, an orphan receptor that is mutated in GCB-DLBCL and another germinal centre B-cell-derived malignancy, Burkitt's lymphoma, also represses germinal centre B-cell growth and promotes confinement via Gα13. These findings identify a Gα13-dependent pathway that exerts dual actions in suppressing growth and blocking dissemination of germinal centre B cells that is frequently disrupted in germinal centre B-cell-derived lymphoma.

We sequenced the *S1PR2* coding region in 117 GCB-DLBCL, 31 Burkitt's lymphoma and 68 activated B-cell-like (ABC)-DLBCL samples. Twelve *S1PR2* coding mutations were identified in the GCB-DLBCL samples versus one in each of the Burkitt's lymphoma and ABC-DLBCL cohorts (Supplementary Tables 1 and 2). The majority of GCB-DLBCL mutations were in conserved transmembrane residues (Fig. 1a) and all were predicted to be structurally damaging. Cell-line transduction experiments showed that five of eight tested mutations disrupted S1PR2 protein expression (Fig. 1b and Extended Data Fig. 1a–c). These same mutations disrupted S1P-mediated inhibition of CXCL12-induced pAkt and migration (Fig. 1c, d). One additional mutant, R147C, which was expressed at levels similar to wild type (WT) (Fig. 1b and Extended Data Fig. 1), showed a strongly reduced ability to support S1P-mediated inhibition

of pAkt and migration (Fig. 1c, d and Extended Data Fig. 1d, e). These observations suggested that tumours harbouring single mutant *S1PR2* alleles (Extended Data Fig. 2) are often likely to be functionally heterozygous for *S1PR2*. Using a mixed bone-marrow chimaera system in mice[3], *S1pr2* heterozygous B cells showed marked expansion in the germinal centre (GC) relative to the follicular compartment in mesenteric lymph nodes and Peyer's patches of unimmunized mice (Fig. 1e and Extended Data Fig. 3a, b). Overexpression of WT S1PR2 repressed the outgrowth of $S1pr2^{+/-}$ GC B cells and this was also seen in mutant R329C, whereas the R147C mutation caused the receptor to lose GC growth suppressive activity (Fig. 1f and Extended Data Fig. 3c, d). On the basis of molecular simulation analysis (Supplementary Information and Extended Data Fig. 3e–g) we propose that the R147C S1PR2 mutant cannot attain the active conformation necessary for G-protein recruitment and signalling.

Gα12 and Gα13 often function redundantly in transmitting G-protein-coupled receptor signals[8]. Transcripts for both G-proteins are upregulated in GC B cells, with *Gna13* transcripts appearing more abundant (Extended Data Fig. 4a). In accord with recent whole-exome sequencing studies that reported mutations in *GNA13* but not *GNA12* (refs 5, 6 and 9–11), we found frequent *GNA13* coding mutations in GCB-DLBCL and Burkitt's lymphoma biopsy samples, with a number of biallelic cases (Supplementary Table 2 and Extended Data Fig. 2). Analysis of mixed bone-marrow chimaeras revealed that Gα13 deficiency was sufficient to confer a GC B-cell growth advantage in mesenteric lymph nodes and to a lesser extent in Peyer's patches (Fig. 1g and Extended Data Fig. 4b). Gα13-deficient mesenteric lymph node GC B cells showed increased pAkt relative to WT when incubated *ex vivo* with CXCL12 and S1P (Fig. 1h). Deficiency in the Gα13 effector, Arhgef1 (p115 RhoGEF or Lsc), led to a similar defect in the ability of S1P to repress chemokine induced pAkt (Fig. 1i).

To determine whether loss of Gα13 in B cells could promote lymphomagenesis, we allowed a cohort of *Gna13*-deficient mice to age. At 1 year, 10 out of 18 *Gna13*-deficient mice showed a greater than tenfold expansion of GC B cells compared with littermate controls (Fig. 1j, k), and at least five of the outgrowths appeared clonal (Extended Data Fig. 4c). Three of the *Gna13*-deficient animals showed massive mesenteric lymphadenopathy (Fig. 1l and data not shown), with evidence in one case (number 307) of spleen and Peyer's patch involvement (Fig. 1l and Extended Data Fig. 4c–e). Immunophenotyping of the Gα13-deficient tumours confirmed they were of GC origin (Extended Data Fig. 4f).

[1]Department of Microbiology and Immunology, University of California, San Francisco, California, 94143, USA. [2]Department of Medicine, University of California, San Francisco, California 94143, USA. [3]Howard Hughes Medical Institute, University of California, San Francisco, California 94143, USA. [4]Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [5]Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA. [6]Department of Pathology, University of Würzburg, 97080 Würzburg, Germany. [7]Department of Clinical Pathology, Robert-Bosch-Krankenhaus, Auerbachstraße 110, 70376 Stuttgart, Germany. [8]Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, 70376 Stuttgart, Germany. [9]British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. [10]Department of Pathology, University of Arizona, Tucson, Arizona 85724, USA. [11]Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain. [12]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [13]Pathology Clinic, Rikshospitalet University Hospital, 0372 Oslo, Norway. [14]Institute for Cancer Research, Rikshospitalet University Hospital, University of Oslo, 0310 Oslo, Norway. [15]Center for Cancer Biomedicine, Faculty Division of the Norwegian Radium Hospital, University of Oslo, 0310 Oslo, Norway. [16]Oregon Health and Science University, Portland, Oregon 97239, USA. [17]Cleveland Clinic Pathology and Laboratory Medicine Institute, Cleveland, Ohio 44195, USA. [18]Department of Pathology, City of Hope National Medical Center, Duarte, California 91010, USA. [19]Department of Pathology, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. [20]Department of Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. †Present address: Immunology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
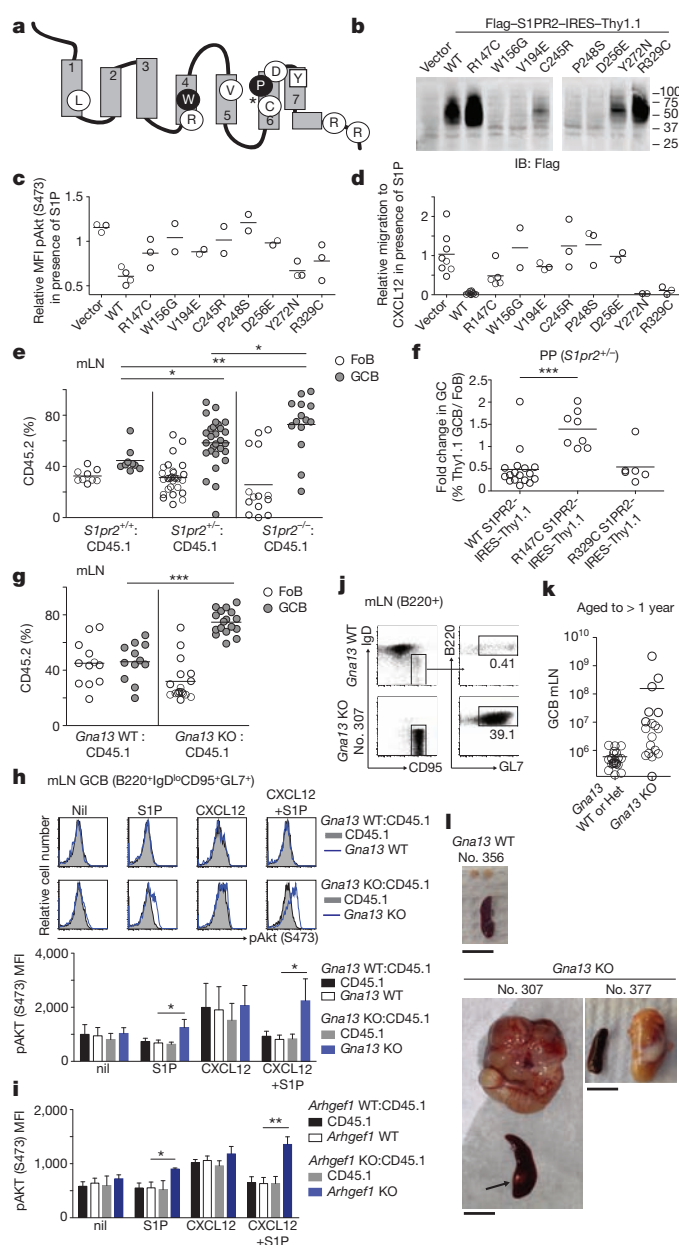
Figure 1 | Lymphoma-associated S1PR2 mutations are functionally disruptive and loss of Gα13 is sufficient to promote GC B-cell survival and lymphomagenesis. a, Schematic of S1PR2 with mutated residues highlighted. Circles denote mutated residues conserved in S1PR2 across species, filled circles denote those conserved across type A G-protein-coupled receptors, squares denote residues not conserved across species, and the asterisk is the position of truncating frameshift mutation. **b**, Western blot of Flag expression in WEHI231 cells transduced with Flag-tagged WT or mutant S1PR2 or empty vector. One experiment representative of three independent biological replicates is shown. The gap in the gel image marks the position of one lane that was not relevant to this experiment and was removed for clarity. **c**, WEHI231 cells transduced as in **b** were stimulated with CXCL12 (100 ng ml^{-1}) in the presence or absence of S1P (1 nM) for 5 min and analysed for phosphorylation of Akt (pAkt S473) by intracellular fluorescence-activated cell sorting (FACS). Mean fluorescence intensity of pAkt in samples treated with both CXCL12 and S1P relative to CXCL12 alone is shown. Data are pooled from four independent experiments. **d**, Transwell migration of cells transduced as in **b**, in response to CXCL12 (100 ng ml^{-1}) in the presence or absence of S1P (1 nM). The relative migration of transduced cells to CXCL12 in the presence versus absence of S1P is shown. Data are pooled from eight independent experiments. **e**, Percentages of CD45.2 follicular B cells (FoB) and GC B cells from mesenteric lymph nodes of mixed bone-marrow chimaeras generated with ~70% WT CD45.1 cells and ~30% $S1pr2$ WT ($n = 9$), heterozygous ($n = 28$) or knockout ($n = 19$) CD45.2 bone marrow, assessed by FACS. Gating scheme is shown in Extended Data Fig. 3a. Data are pooled from four independent experiments. **f**, Fold change in frequency of Thy1.1 reporter$^+$ cells in GC relative to follicular B cells of Peyer's patches from chimaeras reconstituted with $S1pr2^{+/-}$ bone marrow transduced with retrovirus expressing either WT ($n = 17$) or mutant S1PR2 (R147C, $n = 8$; R329C, $n = 6$). Gating scheme is shown in Extended Data Fig. 3c. Data are pooled from three independent experiments. **g**, Percentages of CD45.2$^+$ follicular and GC B cells from mesenteric lymph nodes of mixed bone-marrow chimaeras generated with ~40% $Gna13$ WT ($f/+$) ($n = 12$) or KO (f/f $mb1$-cre) ($n = 17$) CD45.2 cells and ~60% WT CD45.1 cells. Data are pooled from four independent experiments. **h**, **i**, Intracellular FACS for pAkt in GC B cells from mesenteric lymph node of $Gna13$ (**h**) or $Arhgef1$ (**i**) mixed bone-marrow chimaeras that were stimulated *ex vivo* with or without CXCL12 (300 ng ml^{-1}) in the presence or absence of S1P (10 nM) for 10 min. Data are mean ± s.e.m. from one experiment with three biological replicates for each treatment and are representative of four experiments ($Gna13$) or three experiments ($Arhgef1$). **j**, FACS analysis of mesenteric lymph node of 1-year-old $Gna13$ WT or $Gna13$ KO (number 307). Percentage of total cells that are GC B cells is indicated. **k**, GC B-cell number from mesenteric lymph node of $Gna13$ WT and heterozygous ($n = 20$) or KO ($n = 18$) animals aged to 12–16 months. **l**, Gross appearance of mesenteric lymph node and spleen from $Gna13$ control and two $Gna13$ KO animals. Arrow in number 307 denotes splenic nodule (see also Extended Data Fig. 4c–e). Scale bar, 1 cm. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, unpaired two-tailed Student's t-test.

To test the conservation of the Gα13-signalling pathway in human GC B cells, we performed gene rescue experiments in GCB-DLBCL cell lines. Sequencing of *S1PR2*, *GNA13* and *ARHGEF1* in a panel of GCB-DLBCL cell lines identified several with deleterious mutations in these genes (Supplementary Table 3 and Extended Data Fig. 5a). The mutations in *GNA13* matched those previously described and were associated with reduced protein levels[6]. *ARHGEF1* mutations have not previously been reported, probably because the large size (~24 kilobases) of this 27-exon gene and its multiple splice variants and low transcript abundance make sequence analysis difficult. Remarkably, 10 out of 20 cell lines with analysable *ARHGEF1* sequence showed mutations in this gene, several of which resulted in premature stop codons (Supplementary Table 3 and Extended Data Fig. 5a). Using retroviral transduction to restore gene expression, we established that loss of S1PR2, Gα13 and ARHGEF1 were each sufficient to disrupt S1P-mediated suppression of pAKT and, in the case of cell lines that were migratory, to disrupt S1P-mediated inhibition of migration (see Supplementary Information and Extended Data Fig. 5).

The mechanisms by which malignant GC B cells can exit the GC niche and lymphoid organ to spread among multiple lymph nodes or to systemic sites such as bone marrow have not been defined. Consistent with

a lack of migration inhibition by S1P (Fig. 2a), mice lacking Gα13 in B cells showed marked disruption of GC architecture in mesenteric lymph nodes (Fig. 2b and Extended Data Fig. 6a). In a mixed transfer system, Gα13-deficient GC B cells were excluded from the interior of otherwise WT GCs (Extended Data Fig. 6b). Remarkably, Gα13-deficient GC B cells were readily detected in lymph and to a lesser extent in blood while WT GC B cells were absent from circulation (Fig. 2c). In mixed bone-marrow chimaeras, Gα13-deficient GC B cells were again detectable in the lymph, indicating that Gα13 was needed intrinsically in GC B cells to inhibit egress (Fig. 2d). Analysis of Arhgef1-deficient mice and chimaeras revealed a similar disruption of mesenteric lymph node GC architecture (Fig. 2b and Extended Data Fig. 6c) and GC B-cell appearance in lymph and blood (Fig. 2e, f). In contrast, S1PR2-deficient GC B cells were not significantly higher in lymph relative to littermate controls (Fig. 2g). Analysis of mice expressing constitutively active myristoylated Akt (myrAkt) or overexpressing BCL2 in B cells established that increased GC B-cell survival was not sufficient to lead to dissemination (Supplementary Information and Extended Data Fig. 7).

GNA13 mutations and *BCL2* rearrangements and potentially activating mutations frequently occur together in GCB-DLBCL[6,12]. GC cells
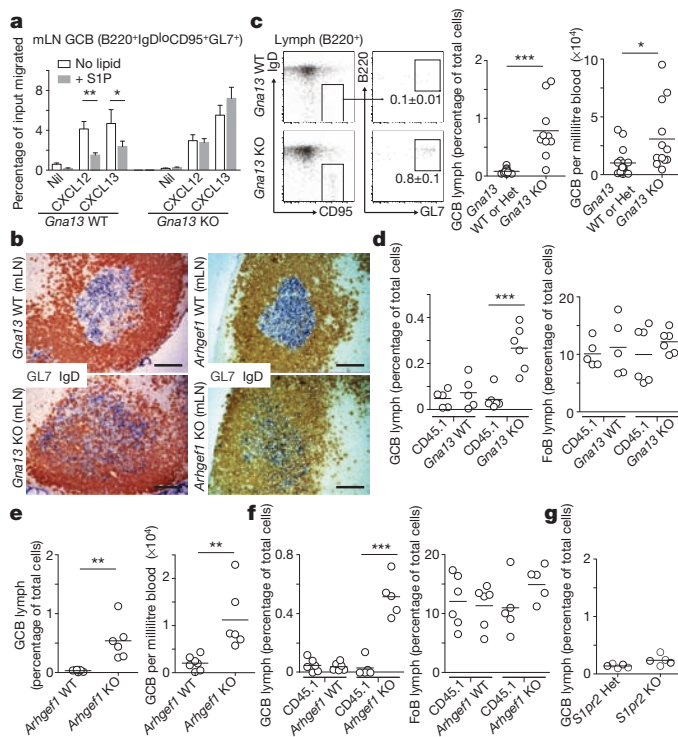
Figure 2 | Loss of confinement and systemic dissemination of GC B cells in the absence of Gα13 or Arhgef1. a, Transwell migration of mesenteric lymph node GC B cells from *Gna13* WT (*f/+*) or *Gna13* KO (*f/f mb1-cre*) mice to CXCL12 (300 ng ml⁻¹) or CXCL13 (1 µg ml⁻¹) in the presence or absence of S1P (10 nM). Data are mean ± s.e.m., pooled from five independent experiments with two technical replicates in each experiment.
b, Immunohistochemical analysis of mesenteric lymph nodes from *Gna13* or *Arhgef1* WT or KO mice stained to detect GC B cells (GL7, blue) and naive follicular B cells (IgD, brown). Scale bar, 100 µm. Data are representative of at least four mesenteric lymph nodes of each type. **c–g**, Lymph and/or blood from *Gna13* WT (*n* = 13 for lymph; *n* = 17 for blood) or KO (*n* = 10 for lymph; *n* = 12 for blood) animals (**c**), *Gna13* mixed chimaeras (WT, *n* = 5; KO, *n* = 6) (**d**), *Arhgef1* WT (*n* = 7) or KO (*n* = 6) animals (**e**), *Arhgef1* mixed chimaeras (WT, *n* = 6; KO, *n* = 5) (**f**) or *S1pr2* heterozygous (*n* = 5) or KO (*n* = 5) animals (**g**) were analysed for the presence of GC B cells by FACS. Representative FACS plot for GC B cells in lymph is shown in **c** with the percentage of total cells that are GC B cells indicated. Data are shown as GC B-cell frequency among total cells in lymph and as cells per millilitre in blood. Data in **c–g** are pooled from between 3 and 13 independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001, unpaired two-tailed Student's *t*-test.

in mice with combined Gα13 deficiency and BCL2 overexpression showed enhanced *ex vivo* survival (Fig. 3a), increased numbers (Fig. 3b), wider dispersal throughout the follicle and interfollicular regions in mesenteric lymph nodes (Extended Data Fig. 7f) and twofold increased frequencies in lymph and blood (compare Figs 3c and 2c), compared with cells in Gα13-deficient mice.

To examine requirements for GC B-cell persistence after arriving at a distant site, we bypassed the egress step and intravenously transferred mesenteric lymph node cells to congenically distinct recipients. Transferred WT GC B cells were essentially undetectable in recipient spleen and bone marrow after 6 hours (Fig. 3d, e) and Gα13 deficiency alone was insufficient to cause a significant increase in their number (Fig. 3e). BCL2-overexpression alone caused an elevation in GC B-cell frequency in recipient spleens but not bone marrow (Fig. 3e). Loss of Gα13 combined with BCL2-overexpression led to greater accumulation of transferred GC B cells in spleen and now led to an increase in their frequency in bone marrow (Fig. 3e). This combinatorial effect probably reflects an ability of Gα13 deficiency and BCL2-overexpression to cooperate in promoting survival of GC B cells outside the GC niche (Fig. 3a). To determine whether GC B cells could seed distant lymph nodes after entry into
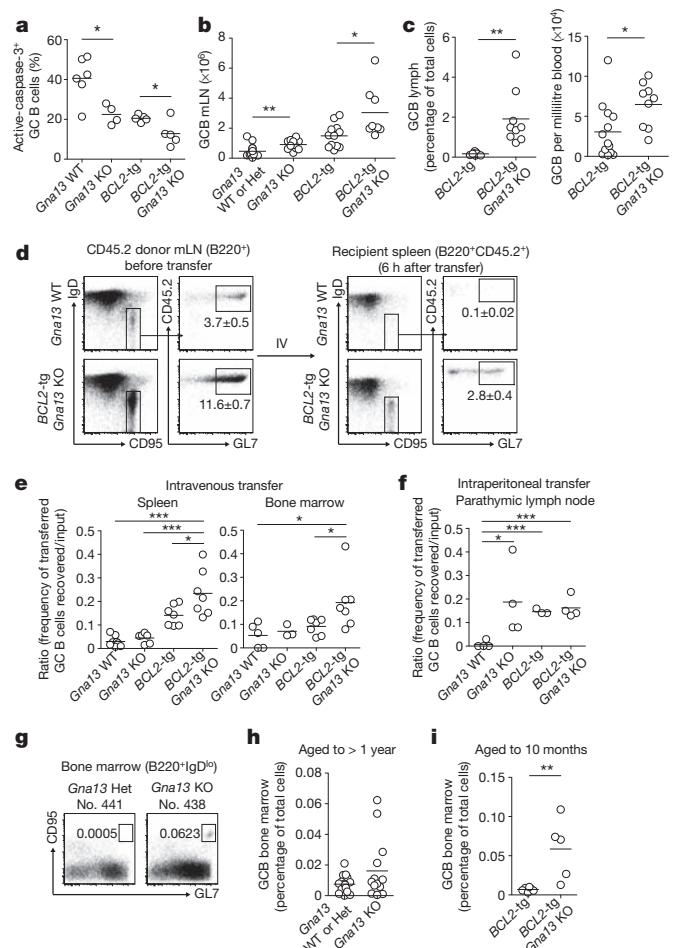
Figure 3 | Gα13 deficiency promotes haematogenous spread and lymphatic seeding of GC B cells in distant organs. a, Intracellular FACS for active caspase-3 in GC B cells from non-*BCL2*-tg or *BCL2*-tg *Gna13* WT or KO mesenteric lymph node cells incubated at 37 °C for 3 h. **b**, GC B-cell numbers in mesenteric lymph nodes from *Gna13* KO or control (*Gna13* WT or Het) mice with or without the *BCL2*-tg, determined by FACS. *n* = 17, 11, 12 and 8, respectively. **c**, GC B cells in lymph and blood of *BCL2*-tg (lymph, *n* = 9; blood, *n* = 13) or *BCL2*-tg *Gna13* KO (lymph, *n* = 9; blood, *n* = 9) mice. **d–f**, mesenteric lymph node cells from non-*BCL2*-tg or *BCL2*-tg *Gna13* WT or KO CD45.2 mice were transferred intravenously (**d** and **e**, *n* = 8, 6, 7 and 7 for spleen, and *n* = 5, 3, 7, 7 for bone marrow, respectively) or intraperitoneally (**f**, *n* = 4, 4, 3 and 4, respectively) into CD45.1 recipients. Spleen and bone marrow (**d** and **e**) or parathymic lymph nodes (**f**) of recipients were harvested after 6 h and analysed for the presence of donor GC B cells. The percentage of donor B cells that were GC B cells is shown in **d**. Ratios of the percentage donor GC B cells recovered from spleen and bone marrow (**e**) or parathymic lymph nodes (**f**) divided by percentage GC B cells in input is shown. **g–i**, Bone marrow of *Gna13* WT and heterozygous (*n* = 16) or KO (*n* = 14) aged to between 12 and 16 months (**g**, **h**) or *BCL2*-tg *Gna13* WT (*n* = 6) or KO (*n* = 5) aged to 10 months (**i**) was analysed for GC B cells by FACS. Data in **a–c**, **e**, **f**, **h** and **i** are pooled from between 3 and 13 independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001, unpaired two-tailed Student's *t*-test.

lymphatics, we transferred mesenteric lymph node cells intraperitoneally. Small numbers of Gα13-deficient, but not WT, GC B cells were detectable in the draining parathymic lymph nodes after 6 hours (Fig. 3f). In this case, recovery of Gα13-deficient GC B cells was not enhanced by the BCL2 transgene. Bone marrow involvement occurs in a fraction of GCB-DLBCL patients and is a predictor of worse disease[13]. In some year-old Gα13-deficient mice showing mesenteric lymph node tumours, GC B cells could be detected in the bone marrow (Fig. 3g, h). Moreover, in aged *BCL2*-tg *Gna13* knockout (KO) but not *BCL2*-tg *Gna13* WT mice, GC B cells were frequently found in the bone marrow (Fig. 3i).
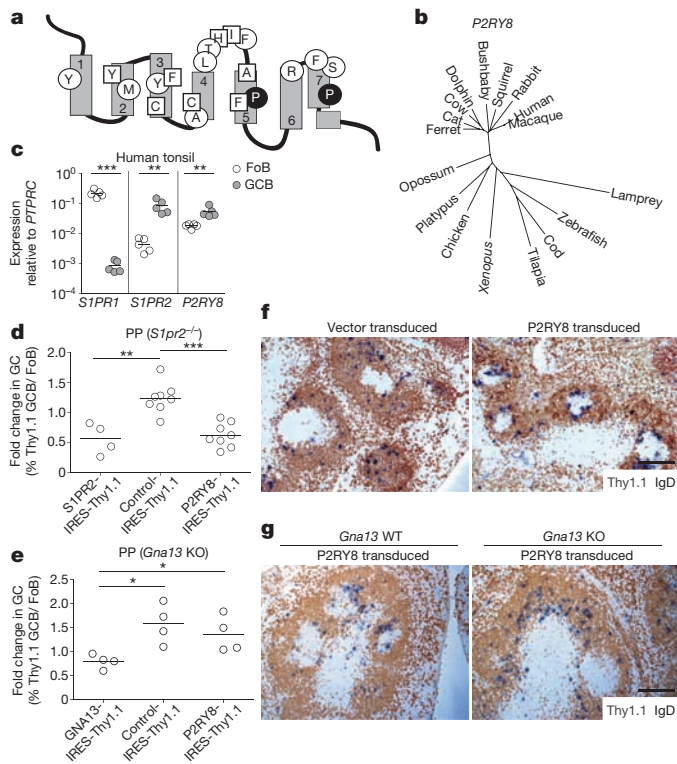
Figure 4 | P2RY8, mutated in GCB-DLBCL and Burkitt's lymphoma, suppresses GC B-cell growth and promotes B-cell confinement via Gα13.
a, Schematic of P2RY8 with locations of mutated residues in GCB-DLBCL and Burkitt's lymphoma. Residues are marked as for S1PR2 in Fig. 1a.
b, Phylogenetic tree of *P2RY8* across species. **c**, Quantitative PCR of *S1PR1*, *S1PR2* and *P2RY8* in FACS-sorted human tonsillar follicular and GC B cells. Data in **c** are from five donors. **d**, **e**, Fold change in frequency of Thy1.1 reporter[+] cells in GC relative to follicular B cells of Peyer's patches from bone marrow chimaeras reconstituted with *S1pr2* KO bone marrow (**d**) or *Gna13* KO (*f/f mb1-cre*) bone marrow (**e**) transduced with retrovirus expressing P2RY8, or with S1PR2, GNA13 or R147C mutant S1PR2 (control). Data in **d** are pooled from two independent experiments (S1PR2, *n* = 4; Control, *n* = 8; P2RY8, *n* = 8). Data in **e** are from one experiment (*n* = 4 in each group).
f, **g**, Immunohistochemical analysis of splenic sections from sheep red blood cell (SRBC)-immunized mice given immunoglobulin (Ig)-transgenic (**f**) or *Gna13* WT or KO (**g**) B cells transduced with retroviral vector encoding Thy1.1 alone (vector) or P2RY8 and Thy1.1, assessed 24 h after cell transfer. Scale bar, 200 μm in **f** and **g**. Data in **f** are representative of three and in **g** of two independent experiments. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, unpaired two-tailed Student's *t*-test.

The more frequent mutations of *GNA13* than of *S1PR2* in both GCB-DLBCL and Burkitt's lymphoma, despite the similar size of their open reading frames, together with our finding of *Gna13*-deficient but not *S1pr2*-deficient mouse GC B cells in circulation (Fig. 2c, g), led us to hypothesize that additional Gα13-coupled G-protein-coupled receptors may be involved in GC B-cell regulation. In this regard, *P2YR8*, a gene situated in the pseudoautosomal region of the X chromosome, was a target of mutations in published whole-exome sequencing data of GCB-DLBCL and Burkitt's lymphoma[5,7,14] and was frequently mutated in our GCB-DLBCL and Burkitt's lymphoma samples, with several of each lymphoma type carrying biallelic mutations (Fig. 4a, Supplementary Table 2 and Extended Data Fig. 2). P2RY8 is an orphan receptor and has orthologues in many vertebrates, but unexpectedly it lacks an orthologue in mouse (Fig. 4b). Like *S1PR2*, *P2RY8* was abundant in human GC B cells (Fig. 4c). Five out of six tested mutations prevented surface P2RY8 expression (Extended Data Fig. 8a, b).

Despite the lack of a mouse *P2RY8* orthologue, we considered the possibility that if the ligand were a small molecule it may be conserved, and we therefore asked whether P2RY8 overexpression influenced GC B-cell

growth. Remarkably, human P2RY8 led to a suppressive effect on GC B-cell growth in mouse Peyer's patches and mesenteric lymph nodes, similar to the effect of S1PR2 overexpression (Fig. 4d and Extended Data Fig. 8c). This suppression required P2RY8 coupling to Gα13 as it was not seen if the cells lacked *Gna13* (Fig. 4e and Extended Data Fig. 8d). In short-term transfers, P2RY8-transduced B cells localized in the centre of the follicle immediately around and often within GCs while vector transduced cells were dispersed throughout the follicle (Fig. 4f, Extended Data Fig. 8e, f and Supplementary Information). In the absence of Gα13, P2RY8 was unable to direct B cells to the follicle centre (Fig. 4g and Extended Data Fig. 8g). Importantly, a control Gα13-coupled G-protein-coupled receptor, Tbxa2r, could not suppress GC B-cell growth or confine cells to the GC niche (Extended Data Fig. 9 and Supplementary Information). These observations lead us to suggest that P2RY8 in humans acts to suppress GC B-cell growth and promote B-cell positioning in a GC location via Gα13-dependent pathways.

GC B cells are normally tightly regulated in their growth and strictly confined to the GC, and they lack the ability to exit into circulation or to survive outside the GC niche. Each of these processes breaks down in the GC B-cell-derived malignancies, GCB-DLBCL and Burkitt's lymphoma. We provide evidence that disruption of Gα13 signalling, via mutations in *GNA13*, *ARHGEF1*, *S1PR2* or *P2RY8*, contributes to this breakdown. *GNA13* is mutated in 15–33% of GCB-DLBCL and ∼15% of Burkitt's lymphoma[6,7,9–11] (Supplementary Table 2 and Extended Data Fig. 2). This is similar to the frequency of mutations in the histone methyltransferases *EZH2* and *MLL2*, deletions of *PTEN* and amplifications of *miR17-92*, genetic alterations that have been highlighted for their role in oncogenesis in GCB-DLBCL[15–21]. Our data support a model (Extended Data Fig. 10 and Supplementary Information) where deleterious mutations in Gα13 and its effector, ARHGEF1, are sufficient to deregulate AKT signalling and to cause loss of confinement, allowing egress of GC B cells into circulation; survival of the disseminating cells at distant sites such as bone marrow depends on co-operating mutations affecting additional genes, such as *BCL2*[12,22]. *S1PR2* and *P2RY8* mutations are also suggested to deregulate AKT signalling and growth but may lead to less dissemination due to overlapping roles in promoting confinement. Potentially inactivating mutations of *RHOA*, a direct target of ARHGEF1[23], have been reported in Burkitt's lymphoma[24]. The mechanism by which RHOA inhibits AKT activation is not yet defined but might involve activation of PTEN or inhibition of RAC[25–27]. We suggest that small molecules that inhibit AKT may replace the missing repressive effects of RHO on growth or survival in cells that harbour defects in the S1PR2/P2RY8–Gα13–ARHGEF1–RHO pathway. Development of active RHO-mimetics may represent a novel therapeutic approach that addresses both lymphoma cell survival and disease dissemination.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Shaffer, A. L., III, Young, R. M. & Staudt, L. M. Pathogenesis of human B cell lymphomas. *Annu. Rev. Immunol.* **30,** 565–610 (2012).
2. Klein, U. & Dalla-Favera, R. Germinal centres: role in B-cell physiology and malignancy. *Nature Rev. Immunol.* **8,** 22–33 (2008).
3. Green, J. A. *et al.* The sphingosine 1-phosphate receptor S1P(2) maintains the homeostasis of germinal center B cells and promotes niche confinement. *Nature Immunol.* **12,** 672–680 (2011).
4. Cattoretti, G. *et al.* Targeted disruption of the S1P2 sphingosine 1-phosphate receptor gene leads to diffuse large B-cell lymphoma formation. *Cancer Res.* **69,** 8686–8692 (2009).
5. Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl Acad. Sci. USA* **109,** 3879–3884 (2012).
6. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476,** 298–303 (2011).
7. Morin, R. D. *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122,** 1256–1265 (2013).
8. Worzfeld, T., Wettschureck, N. & Offermanns, S. G_{12}/G_{13}-mediated signalling in mammalian physiology and disease. *Trends Pharmacol. Sci.* **29,** 582–589 (2008).

9. Love, C. et al. The genetic landscape of mutations in Burkitt lymphoma. Nature Genet. **44,** 1321–1325 (2012).
10. Schmitz, R. et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature **490,** 116–120 (2012).
11. Zhang, J. et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc. Natl Acad. Sci. USA **110,** 1398–1403 (2013).
12. Schuetz, J. M. et al. BCL2 mutations in diffuse large B-cell lymphoma. Leukemia **26,** 1383–1390 (2012).
13. Sehn, L. H. et al. Impact of concordant and discordant bone marrow involvement on outcome in diffuse large B-cell lymphoma treated with R-CHOP. J. Clin. Oncol. **29,** 1452–1457 (2011).
14. Forbes, S. A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. **39,** D945–D950 (2011).
15. Nogai, H., Dorken, B. & Lenz, G. Pathogenesis of non-Hodgkin's lymphoma. J. Clin. Oncol. **29,** 1803–1811 (2011).
16. Morin, R. D. et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nature Genet. **42,** 181–185 (2010).
17. Pfeifer, M. et al. PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma. Proc. Natl Acad. Sci. USA **110,** 12420–12425 (2013).
18. Xiao, C. et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17–92 expression in lymphocytes. Nature Immunol. **9,** 405–414 (2008).
19. Beguelin, W. et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell **23,** 677–692 (2013).
20. Caganova, M. et al. Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. J. Clin. Invest. **123,** 5009–5022 (2013).
21. Pasqualucci, L. et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nature Genet. **43,** 830–837 (2011).
22. Iqbal, J. et al. BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. Am. J. Pathol. **165,** 159–166 (2004).
23. Kozasa, T., Hajicek, N., Chow, C. R. & Suzuki, N. Signalling mechanisms of RhoGTPase regulation by the heterotrimeric G proteins G12 and G13. J. Biochem. **150,** 357–369 (2011).
24. Rohde, M. et al. Recurrent RHOA mutations in pediatric Burkitt lymphoma treated according to the NHL-BFM protocols. Genes Chromosomes Cancer http://dx.doi.org/10.1002/gcc22202 (2014).
25. Li, Z. et al. Regulation of PTEN by Rho small GTPases. Nature Cell Biol. **7,** 399–404 (2005).
26. Michaud, J., Im, D. S. & Hla, T. Inhibitory role of sphingosine 1-phosphate receptor 2 in macrophage recruitment during inflammation. J. Immunol. **184,** 1475–1483 (2010).
27. Guilluy, C., Garcia-Mata, R. & Burridge, K. Rho protein crosstalk: another social network? Trends Cell Biol. **21,** 718–726 (2011).

Supplementary Information is available in the online version of the paper.

Author Contributions J.R.M. designed and performed experiments, interpreted the results and wrote the manuscript. R.S. performed sequencing of cell lines and primary samples, and analysed data. J.A.G. designed experimental procedures used in the manuscript. W.X. analysed sequence data. A.B.L and N.V. performed computer modelling of S1PR2. S.E.B. performed western blots of S1PR2. J.A. performed mouse genotyping and cared for the mouse colony. Y.X. performed quantitative PCR. A.R., G.O., R.D.G., L.M.R., E.C., E.S.J., J.D., E.B.S., R.M.B., R.R.T., J.R.C., D.D.W. and W.C.C. supplied lymphoma patient samples or lines, and reviewed pathological and clinical data. L.M.S. coordinated human sequence analysis, analysed data and supervised research. J.G.C. designed experiments, supervised research and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.G.C. (jason.cyster@ucsf.edu) or L.M.S (lstaudt@mail.nih.gov).

METHODS

Human samples and sequencing. All clinical samples were studied with informed consent according to an institutional review board protocol approved by the National Cancer Institute. Genomic DNA for the single exon coding region of *S1PR2* and complementary DNA (cDNA) for *GNA13* or *ARHGEF1* was amplified by PCR. PCR products were bidirectionally sequenced using an ABI 3730 Genetic Analyzer (Applied Biosystems). Sequence electropherograms were manually reviewed. *ARHGEF1* encodes multiple splice variants with up to 28 coding exons per splice variant. We were unable to sequence the open reading frame of *ARHGEF1* from cDNA in some cell lines in our panel, probably because of splice variation or insufficient transcript. In some cell lines, regions containing coding exons for ARHGEF1 were amplified from genomic DNA. Primers used for amplification and sequencing are shown in Supplementary Table 4. The following NCBI (RefSeq) accession numbers were used to report mutations: *ARHGEF1*, NM_004706 and NP_004697; *GNA13*, NM_006572 and NP_006563; *S1PR2*, NM_004230 and NP_004221.

Mice and bone marrow chimaeras. Adult C57BL6 Ly5.2 (CD45.1$^+$) mice at least 7 weeks of age were from the National Cancer Institute. *S1pr2*$^{-/-}$ mice[28] were backcrossed for at least six generations to C57BL6/J (B6/J). *Arhgef1*$^{-/-}$ mice[29] were backcrossed to B6/J for at least six generations. *Gna13 f/f* mice were on a mixed background[30]. *Mb1-cre* mice (provided by M. Reth) express Cre in all B-lineage cells[31]. *BCL2*-tg mice were of the EµBcl2-22 line[32] that overexpresses BCL2 selectively in B cells. MD4 Ig-tg mice were from an internal colony. Mice lacking *Gna13* in B cells and littermate controls were generated by crossing *mb1-cre + Gna13 f/+* mice to *Gna13 f/f*. In most experiments, bred mice of both sexes were used and were between 7 and 12 weeks of age except in the ageing cohort of *Gna13* animals as indicated. Bone marrow chimaeras were made using Ly5.2 (CD45.1$^+$) from National Cancer Institute as hosts as previously described[33] and analysed at least 8 weeks after reconstitution. For one experiment using *S1pr2* heterozygous and WT littermate donors, mice were also heterozygous for β-2-microglobulin. *CD21-cre* (*Cr2-cre*) mice expressing Cre in mature B cells were from Jackson Laboratory. The mouse genotype was not blinded from the investigator and mice were not randomized. Mice were housed in a specific pathogen-free environment in the Laboratory Animal Research Center at the University of California, San Francisco, and all animal procedures were approved by the Institutional Animal Care and Use Committee.

Retroviral constructs and transductions. S1PR2, P2RY8, GNA13, ARHGEF1 retroviral constructs were made by inserting the human open reading frame into the MSCV2.2 retroviral vector followed by an internal ribosome entry site (IRES) and Thy1.1 or green fluorescent protein (GFP) as an expression marker. The mouse Tbxa2r open reading frame was inserted into the Thy1.1 MSCV2.2 retroviral vector. S1PR2, P2RY8 and Tbxa2r were inserted in frame with a preprolactin leader and Flag-epitope encoding sequence. Lymphoma-associated mutations were introduced into S1PR2 or P2RY8 by quick-change PCR. WEHI231 or human lymphoma cell lines engineered to express an ecotropic retroviral receptor[34] were spin-infected with retrovirus containing vector, WT or mutant S1PR2, P2RY8, Tbxa2r, GNA13 or ARHGEF1. For transduction of bone marrow, *S1pr2* heterozygous or deficient, *CD21-cre* or *Gna13 f/f mb1-cre* donor mice were injected intravenously with 3 mg 5-fluorouracil (Sigma). Bone marrow was collected after 4 days and cultured in DMEM containing 15% (v/v) FBS, antibiotics (penicillin (50 IU ml^{-1}) and streptomycin (50 µg ml^{-1}); Cellgro) and 10 mM HEPES, pH 7.2 (Cellgro), supplemented with IL-3, IL-6 and stem cell factor (at concentrations of 20, 50 or 100 ng ml^{-1}, respectively; Peprotech). Cells were 'spin-infected' twice at days 1 and 2 and transferred into irradiated recipients on day 3. Bone marrow chimaeras in which constitutively active myristoylated Akt (myr-Akt) was selectively expressed in B cells were generated by transducing CD21-cre bone marrow with retrovirus in which myr-Akt was downstream of a loxP–stop–loxP cassette[3]. To generate activated B cells that could be efficiently retrovirally transduced, MD4 Ig-transgenic mice (MGI 2384500) containing lysozyme-specific B cells were injected with 5 mg hen egg lysozyme, splenocytes were harvested 4 h later and the B cells further activated by culturing with 20 µg ml^{-1} anti-CD40 (FGK4.5; BioXcell) for 24 h as in past studies[35]. Alternatively, *Gpr183*$^{+/-}$ or *Gna13* WT or KO spleen cells were harvested in media containing 1 µg ml^{-1} lipopolysaccharide or 0.25 µg ml^{-1} anti-CD180 (RP-105; clone RP14, BD Biosciences) and cultured for 24 h. Later experiments were performed using anti-CD180 activation as we found it much more effective in achieving high levels of transduction than lipopolysaccharide. The activated B cells were spin-infected for 2 h with retroviral supernatant, and cultured overnight before transfer into SRBC-immunized WT mice. Transferred cells were analysed after 24 h by flow cytometry and immunohistochemistry.

Cell isolation, clonality assessment, adoptive transfer, cell culture, treatments, flow cytometry and quantitative PCR. B cells from spleen, mesenteric lymph nodes, Peyer's patches and blood were isolated and stained as previously described[3]. Lymph was collected from the cisterna chyli via fine glass micropipette as previously described[36]. Assessment of clonality by PCR of J558 heavy chain, and κ and λ light chains, from genomic DNA from bulk mesenteric lymph node cells from

year-old mice was performed as previously described[37]. For adoptive transfer experiments, mesenteric lymph nodes were harvested, washed once and transferred intravenously or intraperitoneally into CD45.1 recipient mice. Spleen and bone marrow were harvested 6 h after intravenous transfer; parathymic lymph nodes were harvested 6 h after intraperitoneal transfer. Harvested organs were analysed by FACS for the presence of donor GC B cells. For GC B-cell positioning experiments in a mixed setting, *Gna13* WT or KO CD45.2$^+$ B cells were transferred with WT CD45.1$^+$ B cells into MD4 Ig-tg CD45.1$^+$ recipients. Recipients were then immunized with SRBCs and analysed after 8 days. For pAkt analysis of mesenteric lymph node GC B cells, mesenteric lymph nodes were harvested in RPM-I1640 medium containing 0.5% (w/v) fatty-acid-free BSA (migration media; EMD Biosciences). Cells were RBC lysed twice and re-suspended in migration media. Cells were incubated for 10 min at 37 °C and then stimulated for 10 min with CXCL12 (300 ng ml^{-1}) or S1P (10 nM). Cells were fixed at a final concentration of 1.5% PFA for 10 min at room temperature of 21–23 °C and then permeabilized in ice-cold methanol. Cells were washed twice in staining buffer, blocked with Fc-block (2.4G2; BioXcell) and 5% normal goat serum for 20 min at room temperature of 21–23 °C, stained for 45 min at room temperature of 21–23 °C for Akt phosphorylated at Ser 473 (D9E, number 4060; Cell Signaling Technology) followed by goat antibody to rabbit IgG conjugated to allophycocyanin (sc-3846; Santa Cruz Biotechnology) as well as antibodies to GC markers. For pAkt analysis by flow cytometry in transduced WEHI231 or human GCB-DLBCL lines, cells were stimulated for 5 minutes with or without CXCL12 (100 ng ml^{-1}) with or without S1P (1 nM for WEHI-231 or 10 nM for human GCB-DLBCL lines) and fixed and stained as above for pAkt as well as anti-Thy1.1 conjugated to phycoerythrin (clone ox-7; Biolegend). Human cell lines used in this paper were tested for mycoplasma contamination. Mycoplasma-positive lines were treated with MycoZap (Lonza) and Plasmocin (InvivoGen). All human cell lines were tested for a unique profile of polymorphic DNA copy number variants (CNV fingerprint; unpublished protocol from L. Bergsagel). In some experiments, cells were treated with the PI3K inhibitors wortmannin (Sigma) or GS-1101 (Selleck Chemicals) as negative pAkt staining controls. For active caspase-3 staining, total mesenteric lymph node cells were harvested, washed once and incubated in RPMI-I1640 containing 10% FCS for 3 h at 37 °C; cells were stained for surface markers, fixed and permeabilized with BD Cytofix/Cytoperm and stained with anti-active caspase-3 conjugated to biotin (clone: C92-605; BD Biosciences) according to the manufacturer's instructions. Chemotaxis assays of GC B cells were performed using total mesenteric lymph node cells that were RBC lysed twice or transduced WEHI231 or human GCB-DLBCL lines as described[3]. U-46619 was from Cayman Chemicals. Flow cytometry was performed on a FACSCalibur or LSRii (BD Biosciences). For quantitative PCR analysis of gene expression in GC B cells, *Ptprc* (encoding CD45) was used as a control since its expression was unchanged between follicular and GC B cells by microarray (http://www.immgen.org/ and unpublished data), RNA sequencing analysis (unpublished data) and by surface staining. In contrast, *Gapdh* and *Hprt* were both upregulated in GC B cells (www.Immgen.org and unpublished data).

Western blotting. WEHI231 cells transduced with vector, WT or mutant human S1PR2 were washed twice in migration media and incubated at 37 °C for 30 min, washed once in cold PBS and lysed in 0.5% Brij 35, 0.5% NP40, 150mM NaCl, 10 mM Tris-HCl, pH 7.4 with protease inhibitor cocktail (Roche) for 1 h on ice. Lysates were centrifuged and supernatants were mixed with loading buffer and reducing agent and incubated at room temperature of 21–23 °C for 30 min. Samples were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and Flag expression was detected with rabbit polyclonal anti-Flag (Sigma). For pAkt western blot experiments, Ly7, Ly8 or WEHI cells that were sorted based on Thy1.1 expression and expanded were stimulated as above and lysed in 2× sample buffer, resolved by SDS–PAGE and probed with rabbit anti-pAkt S473 (D9E, number 4060; Cell Signaling Technology).

Immunohistochemical analysis. Cryosections 7 µm in thickness from mesenteric lymph node and spleen were cut and prepared as described[3]. Tumour immunophenotyping was performed using goat polyclonal IRF4 antibody (Santa Cruz, sc-6059) or biotinylated anti-mouse CD138 (clone 281-2; BD Biosciences). For Bcl-6 staining, cryosections were fixed with 4% PFA for 10 min and stained with rabbit polyclonal Bcl6 antibody (Santa Cruz, sc-368). Images were captured with a Zeiss AxioOberver Z1 inverted microscope.

Statistical analysis. Prism software (GraphPad) was used for all statistical analysis. Data were analysed with a two-sample unpaired (or paired, where indicated) Student's *t*-test. *P* values were considered significant when less than 0.05.

28. Kono, M. *et al.* The sphingosine-1-phosphate receptors S1P1, S1P2, and S1P3 function coordinately during embryonic angiogenesis. *J. Biol. Chem.* **279,** 29367–29373 (2004).
29. Francis, S. A., Shen, X., Young, J. B., Kaul, P. & Lerner, D. J. Rho GEF Lsc is required for normal polarization, migration, and adhesion of formyl-peptide-stimulated neutrophils. *Blood* **107,** 1627–1635 (2006).

30. Ruppel, K. M. *et al.* Essential role for Gα13 in endothelial cells during embryonic development. *Proc. Natl Acad. Sci. USA* **102,** 8281–8286 (2005).
31. Hobeika, E. *et al.* Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl Acad. Sci. USA* **103,** 13789–13794 (2006).
32. Strasser, A. *et al.* Enforced bcl2 expression in B-lymphoid cells prolongs antibody responses and elicits autoimmune disease. *Proc. Natl Acad. Sci. USA* **88,** 8661–8665 (1991).
33. Pereira, J. P., Kelly, L. M., Xu, Y. & Cyster, J. G. EBl2 mediates B cell segregation between the outer and centre follicle. *Nature* **460,** 1122–1126 (2009).
34. Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441,** 106–110 (2006).
35. Reif, K. *et al.* Balanced responsiveness to chemoattractants from adjacent zones determines B-cell position. *Nature* **416,** 94–99 (2002).
36. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427,** 355–360 (2004).
37. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449,** 473–477 (2007).
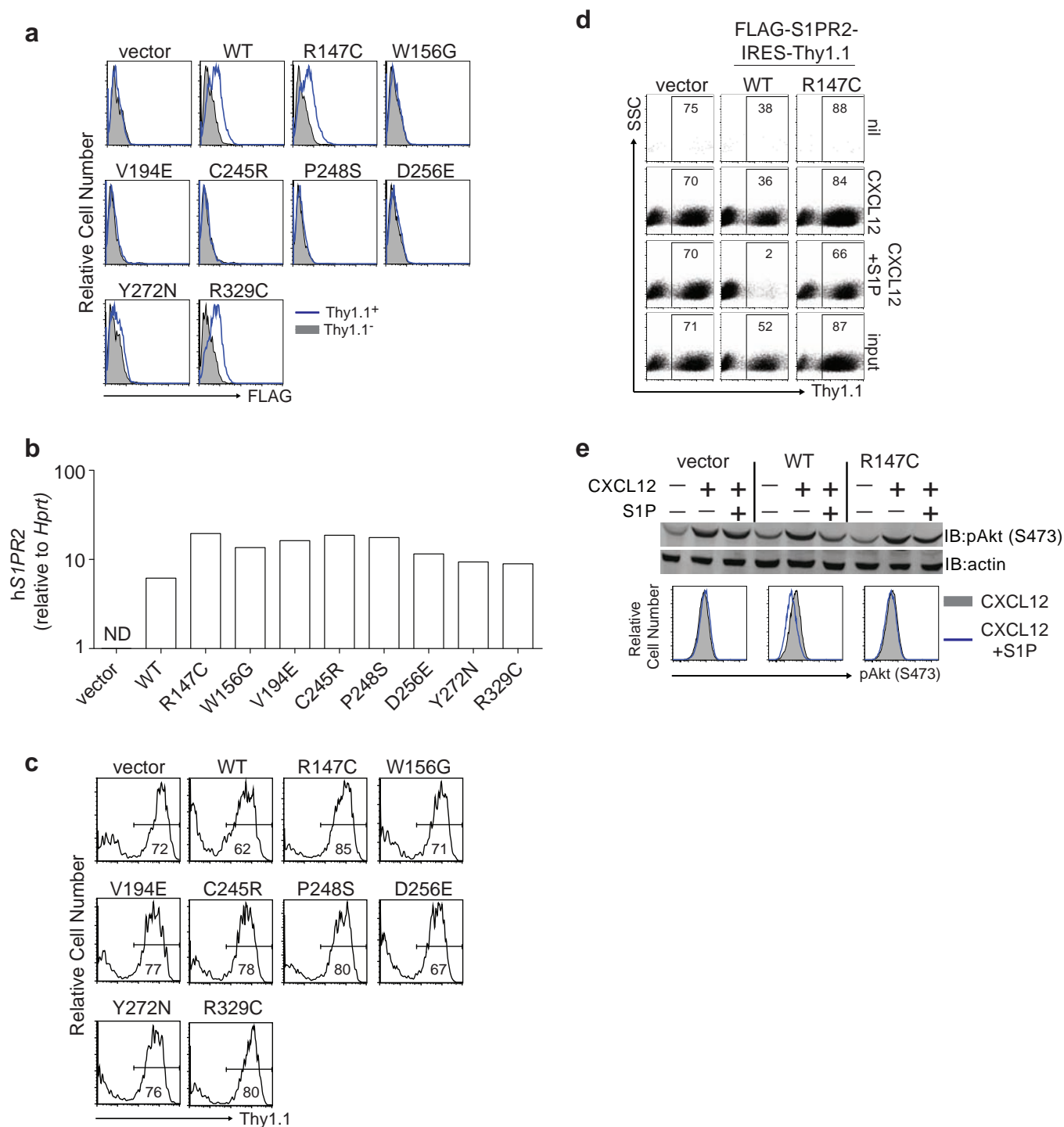
Extended Data Figure 1 | Lymphoma-associated mutations result in loss of expression and function of S1PR2. a–c, Surface expression of Flag (**a**) quantitative PCR of human *S1PR2* (**b**) or Thy1.1 reporter expression (**c**) in mouse WEHI231 B lymphoma cells transduced as described in Fig. 1b. Shown in **a** are histograms of transduced cells (Thy1.1[+]) in blue and untransduced cells (Thy1.1[−]) in grey. Five of eight S1PR2 mutations showed loss of protein expression despite strong transcript and reporter expression. Loss of expression in these five mutants was probably a result of degradation of improperly folded proteins in the endoplasmic reticulum. **d,** Representative FACS plots of transwell migration of WEHI231 cells transduced with vector, WT or R147C mutant S1PR2 to the indicated stimuli or the input sample. Numbers indicate percentage of cells positive for the Thy1.1 reporter. **e,** WEHI231 cells stimulated as in Fig. 1d were analysed for phosphorylation of Akt (pAkt S473) by western blot or by intracellular FACS. Data in **a** and **c** are representative of four independent experiments. Data in **b** are from one experiment. Data in **d** and **e** are representative of three independent experiments.
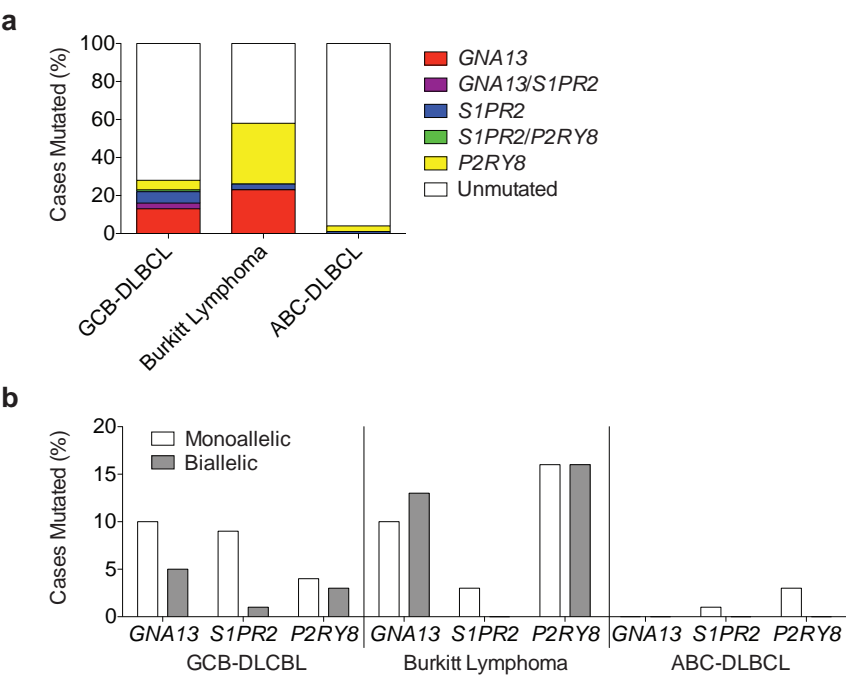
Extended Data Figure 2 | Frequency of mutations in *GNA13*, *S1PR2* and *P2RY8* in aggressive lymphoma. **a**, **b**, Summary of overall mutation frequencies (**a**) and allelic frequencies (**b**) of non-synonymous coding mutations in *S1PR2*, *GNA13* and *P2RY8* in GCB-DLBCL, Burkitt's lymphoma or ABC-DLBCL cases shown in Supplementary Table 2. Unmutated indicates no coding region mutations in the genes shown. Since the sequencing was performed on genomic DNA, the data may underestimate the frequency of biallelic cases as some disruptive mutations may occur in non-coding regulatory elements.

a mLN *S1pr2*+/−(CD45.2) :CD45.1

c PP *S1pr2*+/−

WT S1PR2-IRES-Thy1.1   R147C S1PR2-IRES-Thy1.1

b

d mLN (*S1pr2*+/−)

e
Asp130
Tyr141
TM4
TM3
2.9
Pink: active state
Cyan: inactive state

f
WT ARG147-TYR140
WT TYR140-GLU129
R147C CYS147-TYR140
R147C TYR140-GLU129

g
Wild type
E129
R147
TM4
TM3
2.9
2.8
Y140

R147C mutant
E129
TM4
TM3
C147
Y140
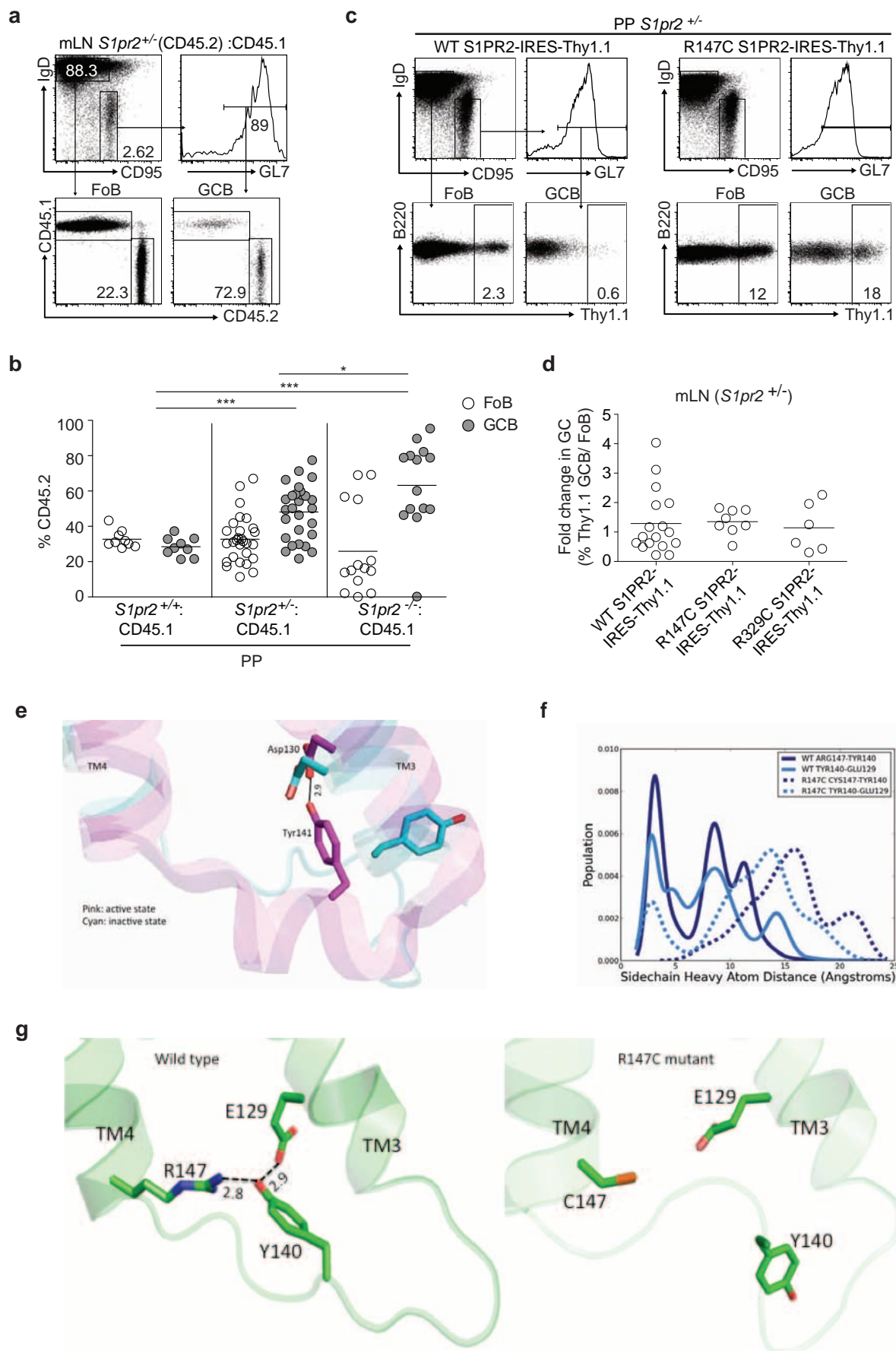
Extended Data Figure 3 | S1PR2 heterozygosity confers a survival advantage to GC B cells and R147C S1PR2 fails to function. a, b, Flow cytometry of follicular and GC B cells from mesenteric lymph node and Peyer's patches of mixed bone-marrow chimaeras generated as in Fig. 1e. Gating strategy for follicular B cells and GCB in mesenteric lymph node is shown in a and percentages of CD45.2$^+$ cells in follicular and GC B cells from Peyer's patches are shown in **b**. Data in **b** are pooled from four independent experiments. **c, d,** Gating strategy of Thy1.1 reporter expression in follicular and GC B cells from Peyer's patches (**c**) or fold change in Thy1.1$^+$ cells in GC relative to follicular B cells of mesenteric lymph node (**d**) of retrovirally transduced bone-marrow chimaeras as described in Fig. 1f. Data in **d** are pooled from three independent experiments. *$P < 0.05$, ***$P < 0.001$, unpaired two-tailed Student's t-test. There was increased variability in mesenteric lymph node relative to Peyer's patches when WT S1PR2 was transduced into $S1PR2^{+/-}$ bone marrow. Nine of 17 animals reconstituted with $S1PR2^{+/-}$ bone marrow

transduced with WT *S1PR2* showed a reduction in expression of Thy1.1 in mesenteric lymph node GC relative to follicular B cells, whereas in six of eight animals reconstituted with R147C *S1PR2* there was increased reporter expression. **e,** The hydrogen bond formed between Y141 in ICL2 and D130 on transmembrane helix 3 (TM3) has been observed only in the active state of β_2-adrenergic receptor (shown in pink) and not in the inactive state (shown in cyan). **f,** Population distribution of the conformational states showing the predicted hydrogen bond network between R147 (TM4), Y140 (ICL2) and E129 (TM3) of the WT (solid lines) and R147C mutant (dashed lines) of S1PR2. **g,** The network of predicted hydrogen bonds mediated by Y140 on ICL2. The hydrogen bond network tightens the interactions between transmembrane helices TM3 and TM4. We hypothesize that this network stabilizes the putative active state conformation of S1PR2. Such a network is broken in the R147C mutant and hence this mutant does not activate the G protein.
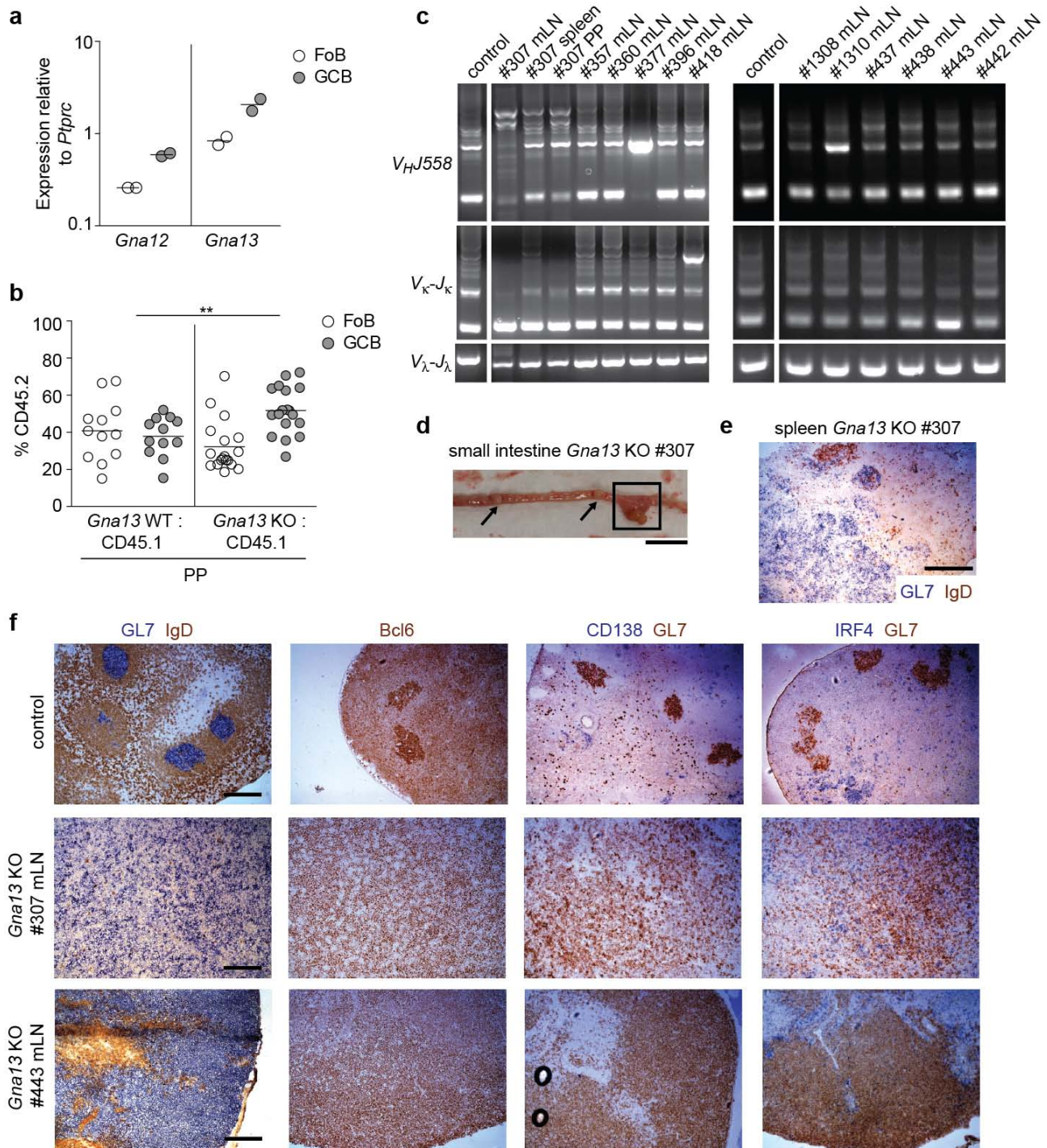
Extended Data Figure 4 | Aged Gα13-deficient mice develop GC-derived lymphoma. **a**, Quantitative PCR analysis of *Gna12* and *Gna13* transcript abundance in follicular and GC B cells relative to the control gene *Ptprc*. **b**, Flow cytometry of follicular and GC B cells from Peyer's patches of mixed bone-marrow chimaeras as described in Fig. 1g. **c**, PCR analysis of V_HJ558–DJ_H, $V_κ$–$J_κ$ and $V_λ$–$J_λ$ rearrangements from indicated tissues of *Gna13* KO animals. The space in the gel image marks the position of lanes that were not relevant to this experiment and were removed for clarity. This PCR analysis was done using bulk rather than sorted GC B cells from tumours and thus probably under-reports the number of animals with clonal outgrowths. Samples scored as having clonal outgrowths (and thus probably harbouring

tumours) were numbers 307, 377, 418, 1310 and 443. In the case of number 307, the splenic nodule and enlarged Peyer's patches showed enrichment of the same VHJ558 clonal bands observed in the mesenteric lymph node. **d**, Gross appearance of small intestine of *Gna13* KO number 307 mouse. Box denotes enlarged Peyer's patches analysed by PCR in **c**; arrows denote two uninvolved Peyer's patches. Scale bar, 1 cm. **e**, Immunohistochemical analysis of splenic nodule from number 307 (see Fig. 1l) for GC marker GL7 (blue) and naive B-cell marker IgD (brown). Scale bar, 500 μm. **f**, Control or enlarged *Gna13* KO mesenteric lymph nodes were stained for the GC B-cell markers GL7 and Bcl6, the plasma cell markers CD138 and IRF4, and the follicular B-cell marker IgD. Scale bar, 200 μm in all samples in **f**.
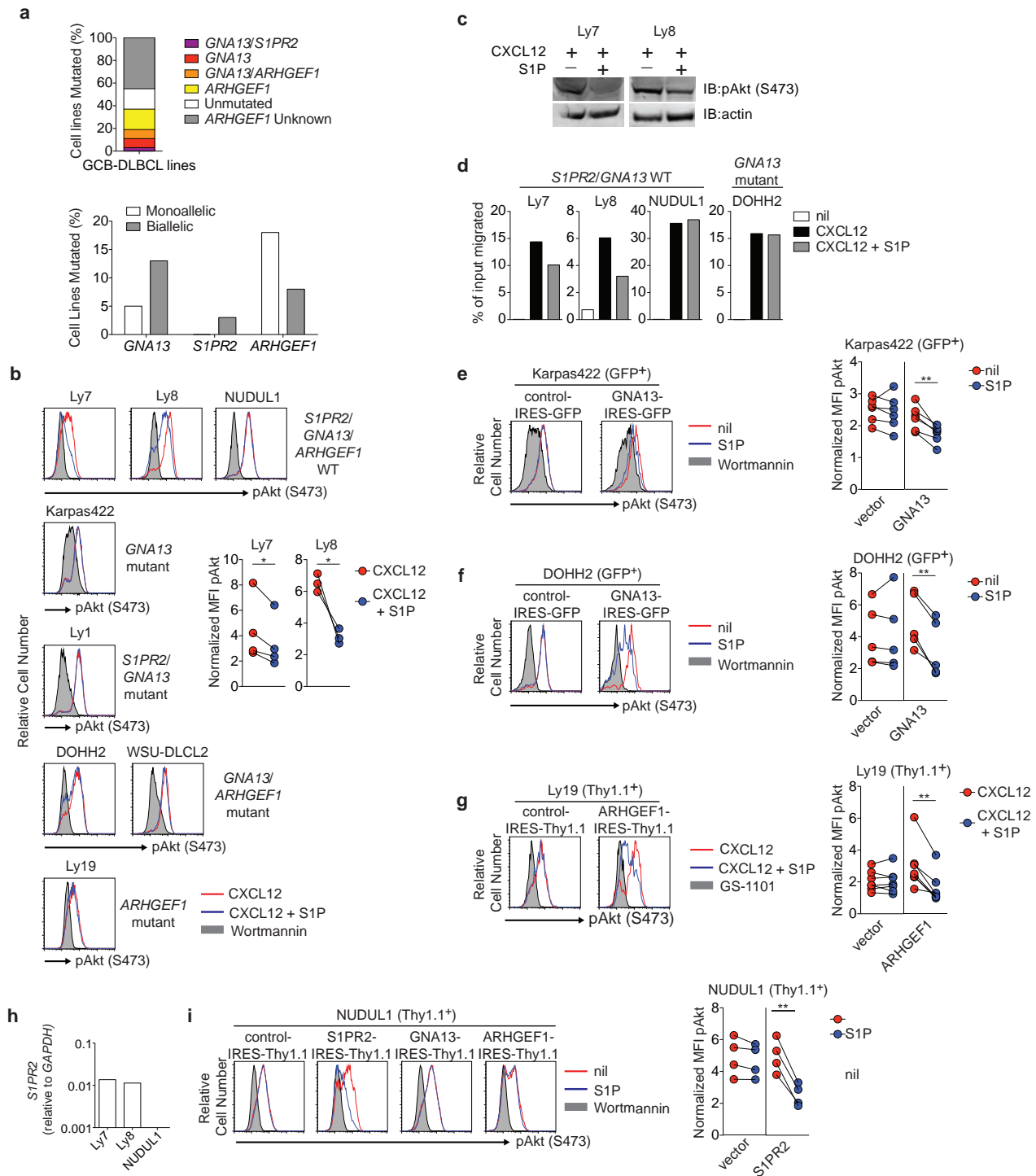
Extended Data Figure 5 | Defective regulation of pAkt and cell migration in human GCB DLBCL cell lines harbouring mutations in the S1PR2 signalling pathway. a, Frequency of non-synonymous coding mutations in *S1PR2, GNA13* and *ARHGEF1* in GCB-DLBCL lines, and the fraction that were mono- or biallelic, summarized from Supplementary Table 3. Unmutated indicates no coding region mutations in the genes shown. **b, c,** Intracellular FACS (**b**) or western blot (**c**) for pAkt in human GCB-DLBCL cell lines that are WT or mutant for *S1PR2, GNA13* or *ARHGEF1* as indicated and which were stimulated with CXCL12 (100 ng ml^{-1}) in the presence or absence of S1P (10 nM) for 5 min. pAkt staining of cells treated with wortmannin (200 nM) for 5 min is shown in grey as a staining control for each cell line. **d,** Transwell migration of *GNA13* WT (Ly7, Ly8, NUDUL1) or mutant (DOHH2) cell lines to CXCL12 (100 ng ml^{-1}) in the presence or absence of S1P (10 nM). **e, f,** Intracellular FACS for pAkt of the *GNA13* mutant cell lines Karpas422

(**d**) or DOHH (**e**) transduced with retrovirus expressing the reporter alone (vector) or GNA13 in the presence or absence of S1P (10 nM) or wortmannin (200 nM; staining control). **g,** Intracellular FACS for pAkt in the *ARHGEF1* mutant cell line Ly19 transduced with retrovirus expressing reporter alone (vector) or ARHGEF1 that were treated as in **b** or with the PI3K inhibitor GS-1101 (2 μM; staining control). **h,** Quantitative PCR analysis of *S1PR2* transcript abundance in human GCB-DLBCL cell lines relative to *GAPDH*. **i,** Intracellular FACS for pAkt in NUDUL1 cells transduced with retrovirus expressing reporter alone (vector), S1PR2, GNA13 or ARHGEF1, treated as in **d**. Data in **b** and **d** are representative of at least three independent experiments. Pooled data from at least three independent experiments are shown in **b, e, f, g** and **i**. Data in **b** are one experiment representative of two. *******P* < 0.01, paired two-tailed Student's *t*-test.
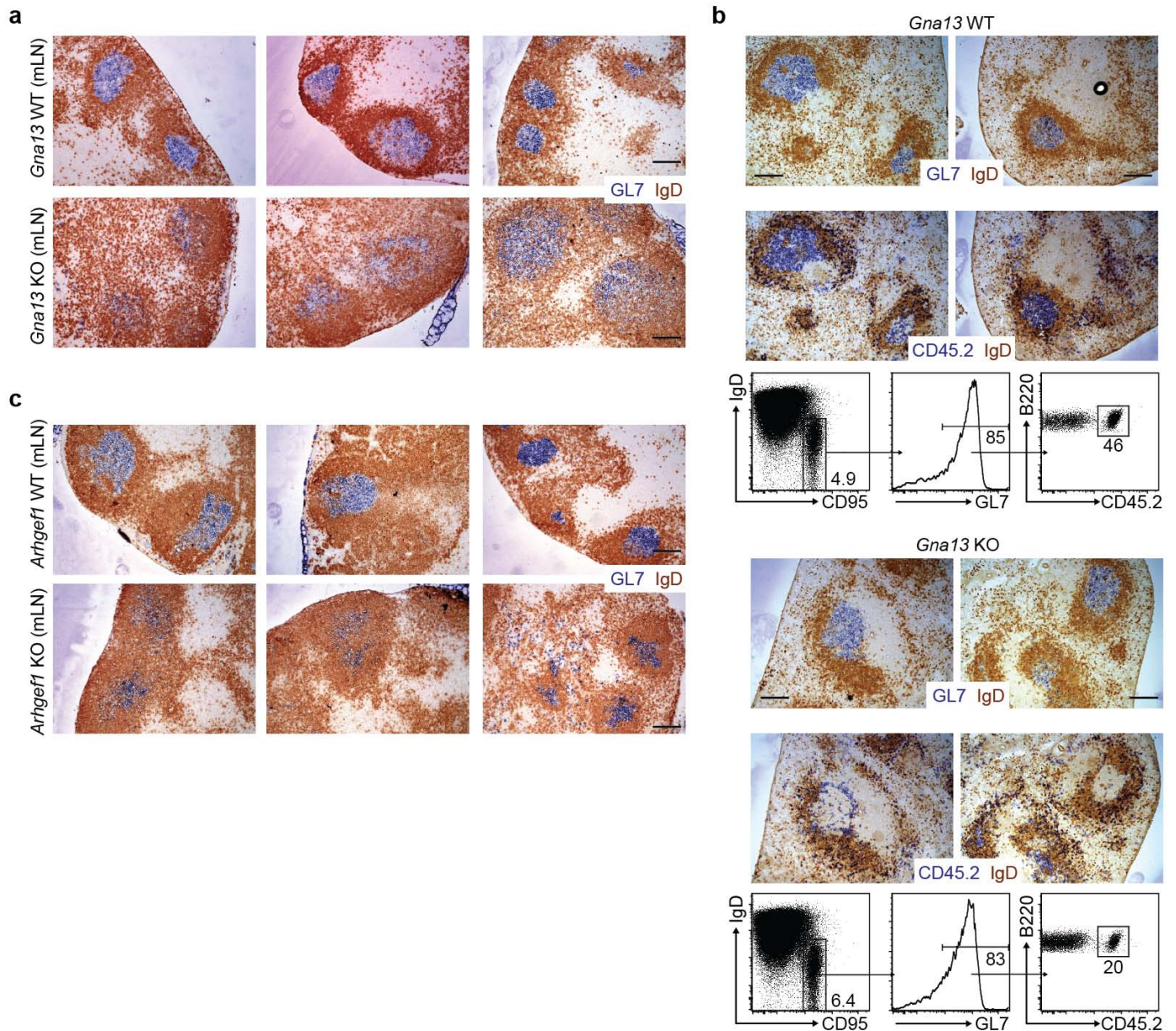
Extended Data Figure 6 | Loss of GC B-cell confinement in the absence of Gα13 or Arhgef1. a, Additional examples of mesenteric lymph node sections from *Gna13* WT or KO mice stained for GC B cells (GL7, blue) and naive B cells (IgD, brown). In the absence of Gα13, the GC border is indistinct and IgD-positive follicular B cells are interspersed with GL7-positive GC B cells throughout the central region of the follicle. The disruption of mesenteric lymph node GC architecture caused by Gα13 deficiency appears more severe than observed in *S1pr2*-deficient mice[3]. b, Mixed B-cell transfer showing exclusion of *Gna13*-deficient GC B cells from the interior of otherwise WT GCs. *Gna13* WT or KO CD45.2[+] B cells were mixed with WT CD45.1[+] B cells and transferred into MD4 Ig-transgenic CD45.1[+] recipients that were then immunized with SRBCs intraperitoneally, and splenic tissue was analysed by immunohistochemistry and FACS after 8 days for CD45.2[+] B cells. This transfer approach allows efficient participation of transferred polyclonal B cells in the GC as the Ig-transgenic recipient B cells are hen-egg lysozyme specific and do not respond to SRBCs. Note that CD45.2[+] WT B cells are distributed uniformly through the GL7[+] GCs (upper panels) whereas the CD45.2[+] *Gna13* KO B cells are located at the perimeter of the GC or in the surrounding follicle (lower panels). In each case, two example images are shown and the GL7 and CD45.2 stains are of adjacent sections. c, Additional sections of mesenteric lymph nodes from *Arhgef1* WT or KO mice, stained for GL7 and IgD. Scale bar, 200 μm in a–c. Data in b are one experiment representative of two.
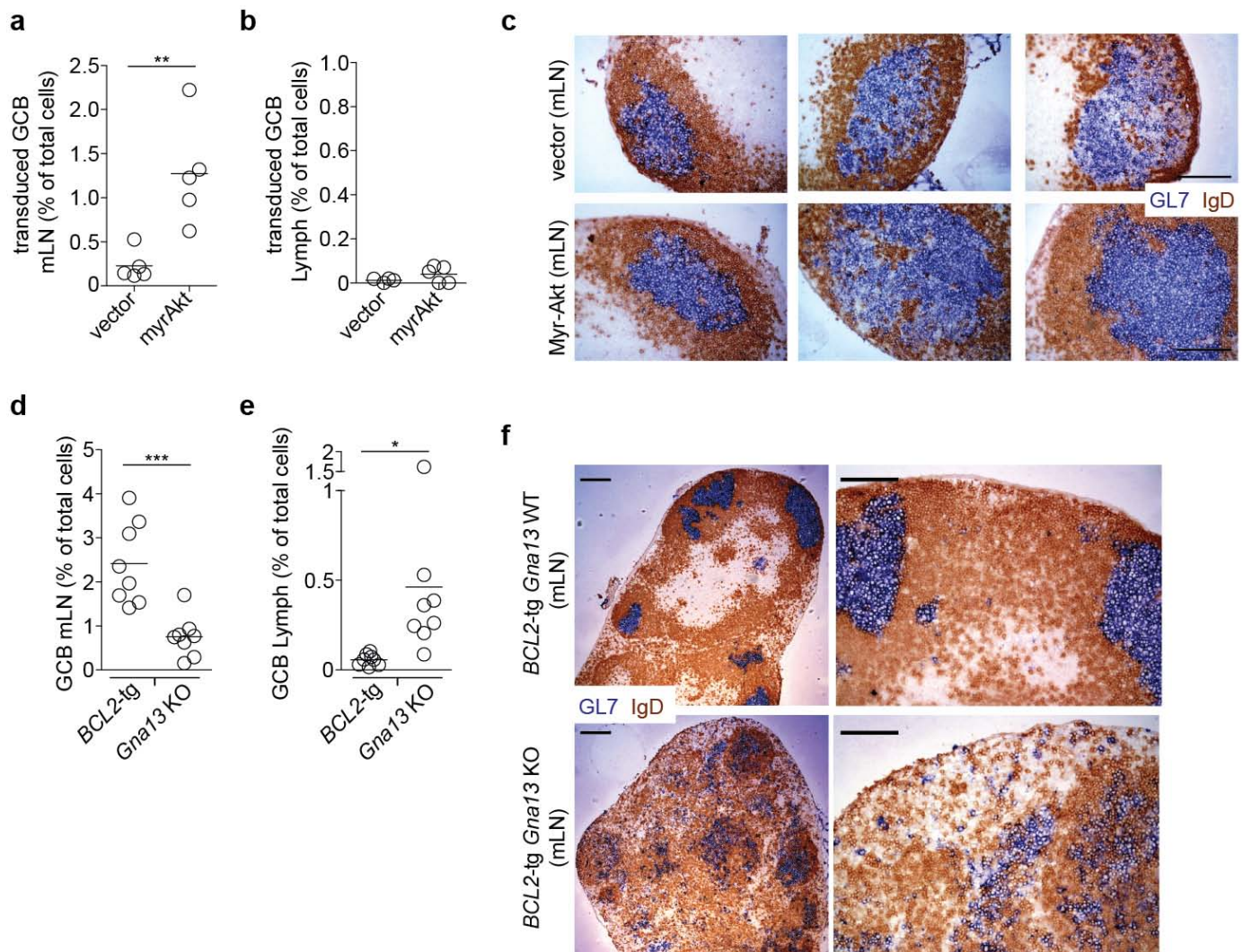
Extended Data Figure 7 | Augmented GC B-cell survival is not sufficient to promote dissemination of GC B cells. a, b, Transduced GC B-cell frequency among total cells in mesenteric lymph node (**a**) and lymph (**b**) of mice reconstituted with bone marrow transduced with B-cell-restricted control (vector, $n = 5$) or myr-Akt ($n = 5$) expressing retrovirus. **c,** Immunohistochemical analysis of mesenteric lymph node sections from mice in **a,** stained for GL7 and IgD. Scale bar, 100 μm. **d, e,** BCL2-tg or *Gna13* KO GC B-cell frequency among total cells in mesenteric lymph node (**d**) and lymph (**e**) of BCL2-tg:*Gna13* KO mixed chimaeras ($n = 8$). **f,** Immunohistochemical analysis of mesenteric lymph nodes from *BCL2*-tg *Gna13* WT or *BCL2*-tg *Gna13* KO mice. Scale bar in low-magnification images (left) is 200 μm and in high-magnification images (right) is 100 μm. Data in **a, b, d** and **e** are pooled from two independent experiments. Data in **c** and **f** are representative of at least three mice of each type.
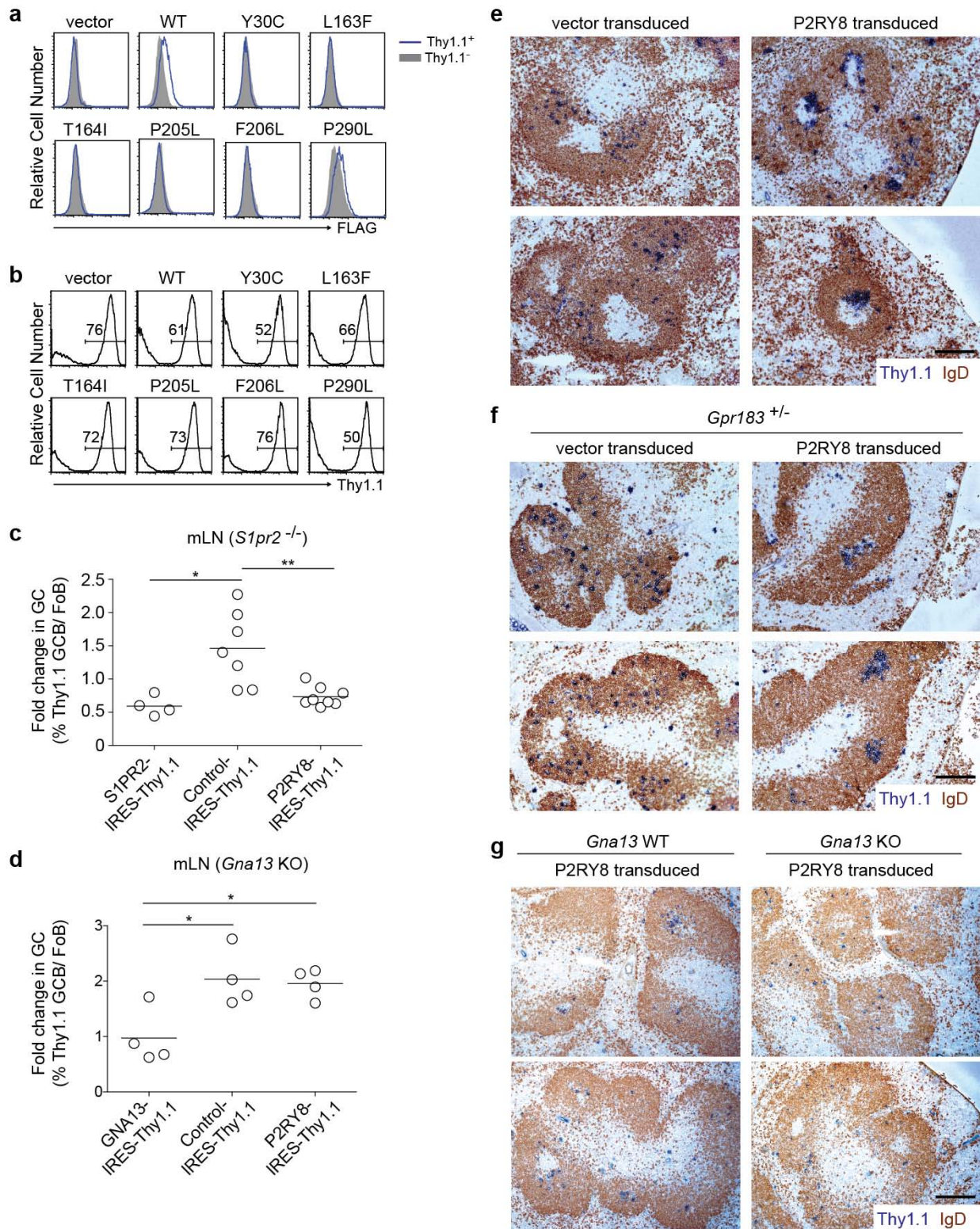
Extended Data Figure 8 | Human P2RY8 suppresses GC B-cell growth and promotes B-cell confinement to the GC in mice. a, b, P2RY8 mutations arising in GCB-DLBCL and Burkitt's lymphoma disrupt receptor expression. Flag-tagged versions of six point mutants and the WT receptor were expressed in WEH231 B cells and surface expression examined by Flag flow cytometry (**a**). The transduction efficiency of each construct was confirmed to be similar based on IRES-Thy1.1 reporter expression (**b**). **c, d,** Fold change in Thy1.1 reporter[+] GC relative to follicular B cells from mesenteric lymph node of chimaeras described in Fig. 4d, e. **e–g,** Immunohistochemical analysis of splenic sections from SRBC-immunized mice given Ig-transgenic (**e**), $Gpr183^{+/-}$ (**f**) or $Gna13$ WT or KO (**g**) B cells transduced as in Fig. 4f, g and assessed 24 h after cell transfer. Data in **a** and **b** are representative of three independent experiments. Data in **e** and **g** are additional examples of the experiments shown in Fig. 4f, g, respectively. Data in **f** are representative of four independent experiments. Scale bar, 200 μm in **e–g**. *$P < 0.05$, **$P < 0.01$, unpaired two-tailed Student's t-test.
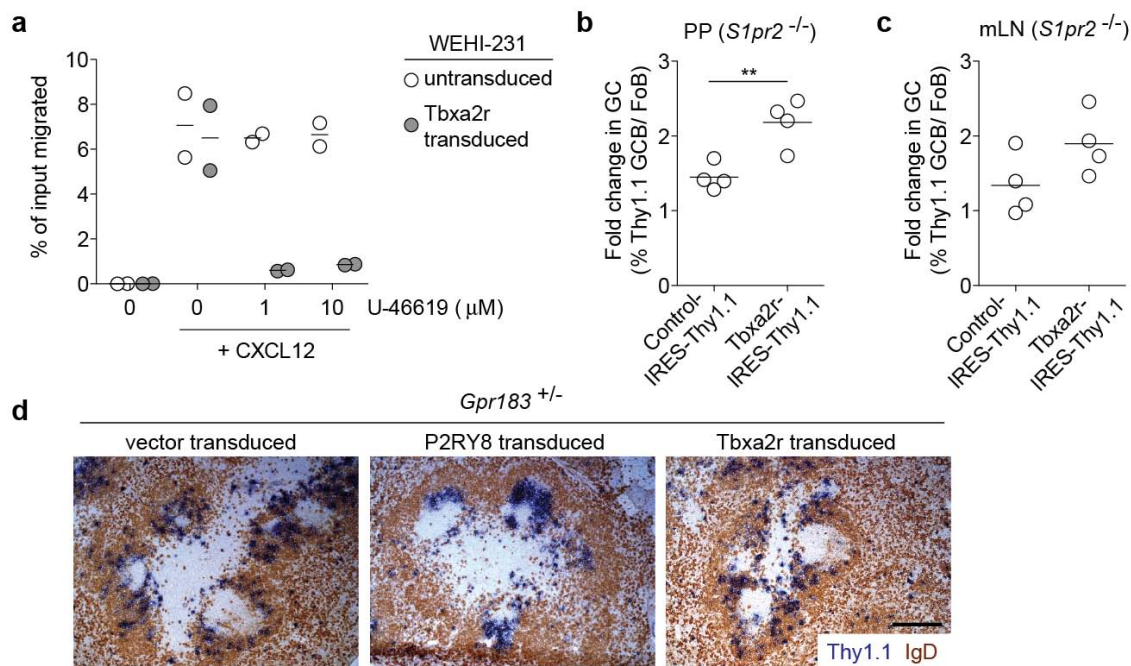
Extended Data Figure 9 | P2RY8-dependent suppression of GC B-cell survival and promotion of B-cell confinement to the GC niche is receptor specific. a, Transwell migration of WEHI231 cells transduced with retrovirus encoding the control Gα13-coupled receptor, Tbxa2r, towards CXCL12 (100 ng ml^{-1}) in the presence or absence of the thromboxane A2 analogue, U-46619. **b, c**, Fold change in frequency of Thy1.1 reporter$^+$ GC relative to follicular B cells of Peyer's patches (**b**) or mesenteric lymph node (**c**) from bone marrow chimaeras reconstituted with *S1pr2* KO bone marrow transduced with empty vector (control) or Tbxa2r. **d**, Immunohistochemical analysis of splenic sections from SRBC-immunized mice given *Gpr183*$^{+/-}$ B cells transduced with empty vector, Tbxa2r or P2RY8, and assessed 24 h after cell transfer. Scale bar, 200 μm. Data in **a** and **d** are one experiment representative of two. Data in **b** and **c** are from one experiment ($n = 4$ in each group). **$P < 0.01$, unpaired two-tailed Student's t-test.

a

S1PR2* & P2RY8*‡

↓

GNA13*‡

↓

ARHGEF1*

↓

Rho(A)‡

↙   ↘

Migration   Akt

*Mutated in GCB-DLBCL
‡Mutated in BL
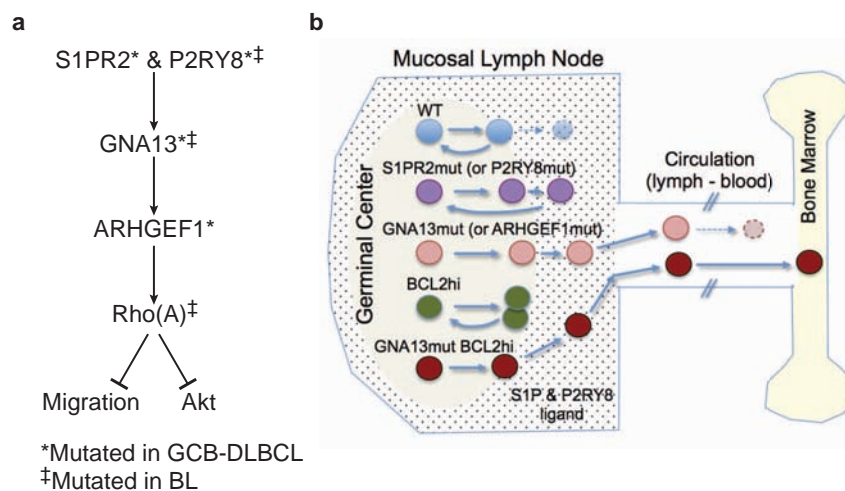
Extended Data Figure 10 | Model relating disruptions in S1PR2/P2RY8–Gα13–ARHGEF1 migration- and Akt-inhibitory pathway to increases in GC B-cell survival, dispersal in the follicle, egress into circulation and dissemination to bone marrow. a, Summary of signalling pathway. **b**, Schematic diagram showing GC-containing lymph node follicle, with connection to efferent lymphatic, blood and bone marrow. Suggested distribution of S1P and of putative P2RY8 ligand within lymph node is shown by dots. Comparative migration and survival behaviour of GC B cells with loss (S1PR2, P2RY8, GNA13, ARHGEF1) or gain (BCL2) of function mutations is summarized.

LETTER

MapZ marks the division sites and positions FtsZ rings in *Streptococcus pneumoniae*

Aurore Fleurie[1]*, Christian Lesterlin[2]*, Sylvie Manuse[1]*, Chao Zhao[1]†, Caroline Cluzel[3], Jean-Pierre Lavergne[1], Mirita Franz-Wachtel[4], Boris Macek[4], Christophe Combet[1], Erkin Kuru[5], Michael S. VanNieuwenhze[5], Yves V. Brun[5], David Sherratt[2] & Christophe Grangeasse[1]

In every living organism, cell division requires accurate identification of the division site and placement of the division machinery. In bacteria, this process is traditionally considered to begin with the polymerization of the highly conserved tubulin-like protein FtsZ into a ring that locates precisely at mid-cell[1]. Over the past decades, several systems have been reported to regulate the spatiotemporal assembly and placement of the FtsZ ring[2–5]. However, the human pathogen *Streptococcus pneumoniae*, in common with many other organisms, is devoid of these canonical systems and the mechanisms of positioning the division machinery remain unknown[4,6]. Here we characterize a novel factor that locates at the division site before FtsZ and guides septum positioning in pneumococcus. Mid-cell-anchored protein Z (MapZ) forms ring structures at the cell equator and moves apart as the cell elongates, therefore behaving as a permanent beacon of division sites. MapZ then positions the FtsZ ring through direct protein–protein interactions. MapZ-mediated control differs from previously described systems mostly on the basis of negative regulation of FtsZ assembly. Furthermore, MapZ is an endogenous target of the Ser/Thr kinase StkP, which was recently shown to have a central role in cytokinesis and morphogenesis of *S. pneumoniae*[7–9]. We show that both phosphorylated and non-phosphorylated forms of MapZ are required for proper Z-ring formation and dynamics. Altogether, this work uncovers a new mechanism for bacterial cell division that is regulated by phosphorylation and illustrates that nature has evolved a diversity of cell division mechanisms adapted to the different bacterial clades.

Recently, some membrane Hanks-type Ser/Thr kinases[10,11] were shown to have a key role in bacterial cell division and morphogenesis[12]. *S. pneumoniae* StkP kinase is crucial for septum assembly, cell shape and localization of peptidoglycan (PG) synthesis[7–9]. However, the underlying regulatory mechanisms by which StkP exerts its function remain elusive. Here, we uncover the role of one endogenous target of StkP, Spr0334, a membrane protein of unknown function that shares no sequence similarity with other known proteins[13]. We named the protein MapZ for mid-cell-anchored protein Z, on the basis of the observations we report here.

In pneumococcus, the *mapZ*-null mutant (Δ*mapZ*) exhibits a variety of aberrant cell shapes and sizes that contrast with wild-type cell morphology (Fig. 1a and Extended Data Fig. 1a). Misshaped Δ*mapZ* cells have mispositioned division septa and form grape-like clusters as observed by electron microscopy (Fig. 1b). These phenotypes are associated with growth defects, as indicated by a 48% increase in generation time and a 30% decrease in cell viability (Extended Data Table 1). Normal cell shape, viability and growth were restored when *mapZ* was reinserted into the chromosome (*mapZ*⁺) or complemented from an ectopic chromosome locus (Δ*mapZ*/P$_{Zn}$-*mapZ*) (Extended Data Fig. 1 and Extended Data Table 1). Bioinformatics analysis[14] of the MapZ sequence predicted the presence of a single transmembrane segment separating a cytoplasmic amino-terminal domain and an extracellular carboxy-terminal domain

(Fig. 1c and Extended Data Fig. 2a). Wide-field microscopy showed that MapZ fused to green fluorescent protein (GFP–MapZ; N-terminal fusion) forms rings positioned at mid-cell and at future division sites (Fig. 1c). Consistent with domain prediction, no fluorescence was detected for MapZ–GFP (extracellular C-terminal fusion) (Extended Data Fig. 2b). The C-terminal extracellular domain is required for MapZ septal localization, as its deletion results in redistribution of MapZ all over the cell
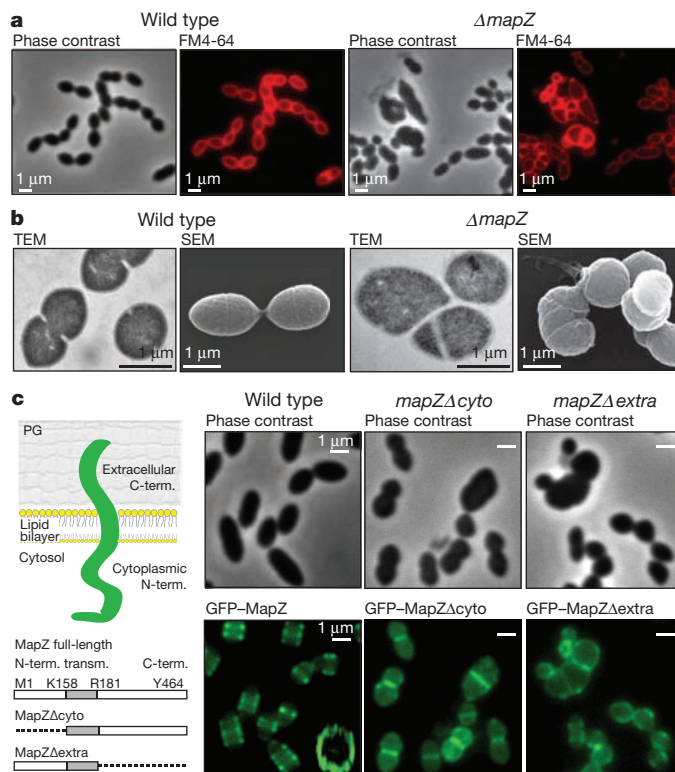
Figure 1 | Characterization of MapZ. a, Cell shape was observed by phase contrast microscopy and after membrane staining with FM4-64. **b,** Transmission electron microscopy (TEM) and scanning electron microscopy (SEM). **c,** Diagram of MapZ domain prediction using TOPCONS (http://topcons.net; see Methods), with an intracellular N-terminal domain (N-term.), a transmembrane domain (Transm.) and an extracellular C-terminal domain (C-term.). Wide-field microscopy images show the localization of GFP–MapZ full-length and GFP–MapZΔcyto and GFP–MapZΔextra, in which the N-terminal or the C-terminal domain are deleted, respectively. Images are representative of experiments made in triplicate.

[1]Bases Moléculaires et Structurales des Systèmes Infectieux, IBCP, Université Lyon 1, CNRS, UMR 5086, Lyon 69007, France. [2]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [3]Laboratoire de Biologie Tissulaire et d'Ingénierie Thrérapeutique, IBCP, Université Lyon 1, CNRS, UMR 5305, Lyon 69007, France. [4]Proteome Center Tübingen, University of Tübingen, Auf der Morgenstelle 15, Tübingen 72076, Germany. [5]Departments of Biology and Chemistry, Indiana University, Bloomington, Indiana 47405, USA. †Present address: Key laboratory of Medical Molecular Virology, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai 200032, China.
*These authors contributed equally to this work.
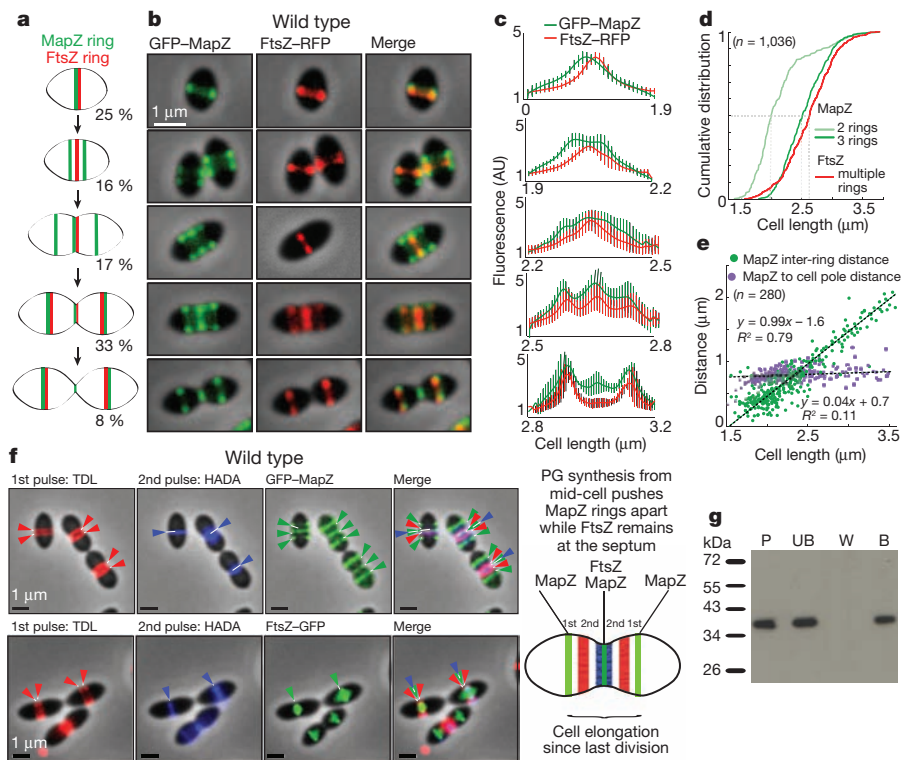
Figure 2 | Localization of MapZ and FtsZ in wild-type cells. a, MapZ- and FtsZ-ring positions during growth (with corresponding cell ratios) **b**, Microscopy of GFP–MapZ and FtsZ fused to red fluorescent protein (FtsZ–RFP). **c**, Fluorescence intensities for different cell size categories (error bars show standard deviation (s.d.) for 10 cells analysed). AU, arbitrary units. **d**, Cumulative distribution of cells with MapZ or FtsZ rings. Dashed lines show cumulative distribution of 0.5. **e**, Distance between outer MapZ rings and between MapZ rings and the closest pole. Linear fitting curve with equation and R^2 values are shown. **f**, Localization of consecutive PG incorporation, TDL (a fluorescent carboxytetramethylrhodamine derivative of D-alanine) (red) and HADA (a fluorescent hydroxy coumarin derivative of D-alanine) (blue), together with GFP–MapZ or FtsZ–GFP. Summary diagram is presented. **g**, Interaction of MapZ extracellular domain with the cell wall (sample of n cells analysed). Images are representative of experiments made in triplicate. B, MapZ bound; P, purified MapZ alone; UB, unbound; W, wash fraction.

membrane, whereas GFP–MapZΔcyto mostly retained septum localization (Fig. 1c). However, both the intra- and the extracellular domains are required for MapZ cellular function because *mapZΔextra* and *mapZΔcyto* strains exhibit morphological and growth defects similar to Δ*mapZ* (Fig. 1c, Extended Data Fig. 1 and Extended Data Table 1).

In newborn wild-type cells, the MapZ ring and FtsZ ring colocalize at mid-cell (Fig. 2a–c). As cells begin elongating, the MapZ ring splits into two rings whereas a single FtsZ ring remains at mid-cell. At an average

cell size of 1.95 µm, 50% of cells have two MapZ rings while >90% of cells still have one single FtsZ ring (Fig. 2d). The early and progressive splitting of MapZ was observed in detail by time-lapse (Extended Data Fig. 4a and Supplementary Video 1) and three-dimensional structured illumination (3D-SIM) snapshot microscopy (Extended Data Fig. 4b–d). At a later stage (average cell size of 2.5 µm), the appearance of a third MapZ ring at mid-cell is shortly followed by splitting of FtsZ into rings that migrate apart until colocalization with MapZ outer rings at future
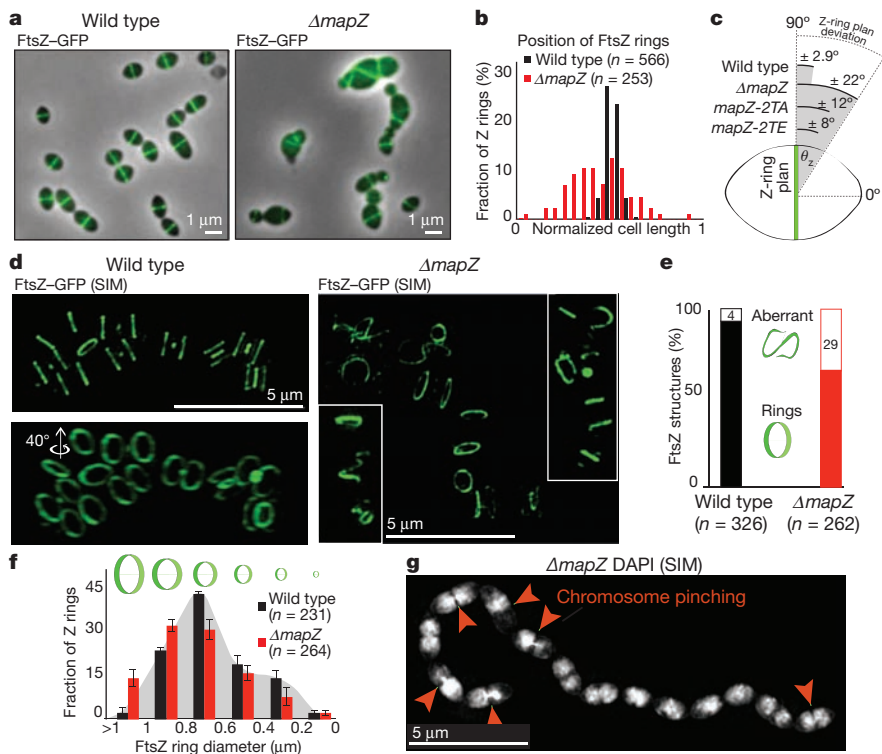
Figure 3 | FtsZ localization in wild-type and Δ*mapZ* cells. a, Localization of FtsZ–GFP. **b**, Positioning of single FtsZ rings. **c**, Deviation of FtsZ-ring angles (θ_z) (60 cells analysed for each strain; P value $< 2.9 \times 10^{-2}$). **d**, 3D-SIM of FtsZ rings. **e**, Fraction of FtsZ rings or aberrant structures (P value $= 1.4 \times 10^{-10}$, two-tailed Student's t-test). **f**, FtsZ-ring diameter distribution (error bars show s.d. from three independent experiments; P value $= 1.1 \times 10^{-3}$). **g**, 3D-SIM of Δ*mapZ* cells after DNA staining with DAPI (sample of n cells analysed). Images are representative of experiments made in triplicate.

division sites (Fig. 2a–d and Extended Data Fig. 3a, b). The mid-cell MapZ/FtsZ ring constricts and eventually closes (Fig. 2a–c and Extended Data Fig. 3a–c) to complete cytokinesis. Identical MapZ and FtsZ localizations were observed in the *rfp-mapZ ftsZ-gfp* strain (Extended Data Fig. 3e).

Importantly, the distance between the outer MapZ rings increases linearly as a function of the cell length, whereas the distance between MapZ and the cell pole remains constant (Fig. 2e). This suggested that MapZ rings are permanently associated with the future division sites and are mechanically pushed apart as PG synthesis forms the new cell halves. This hypothesis was confirmed by visualizing GFP–MapZ together with PG synthesis using sequential incorporation of two fluorescently labelled D-amino acids[15] (Methods). Consistent with a previous report[16], our results showed that PG is incorporated at mid-cell (Fig. 2f). The last synthetized PG (Fig. 2f, blue) pushes the previously incorporated one (Fig. 2f, red) and both are flanked by MapZ rings, while a single FtsZ ring is present at mid-cell (Fig. 2f). The dependence of MapZ positioning on PG synthesis is further supported by the observation that the extracellular C-terminal domain of MapZ efficiently binds PG (Fig. 2g) and that specific inhibition of PG synthesis using vancomycin led to rapid delocalization of MapZ (Extended Data Fig. 5a).

Δ*mapZ* (Fig. 3a, b), *mapZΔextra* and *mapZΔcyto* strains (Extended Data Fig. 5b–e) exhibited severe alterations of FtsZ ring morphology and localization. FtsZ is unable to position at mid-cell (Fig. 3b) and the angles of Z rings with respect to the cell long axis (θ_z) are incorrect, reflecting the inability of FtsZ to find the orthogonal division plan (Fig. 3c). 3D-SIM further revealed major defects of FtsZ structures in the Δ*mapZ* strain (Fig. 3d and compare Supplementary Videos 2 and 3). Aberrant 'non-ring' FtsZ structures are observed in 29% of Δ*mapZ* cells (Fig. 3e). Time-lapse microscopy revealed that FtsZ forms polymers that fail to position correctly and degenerate into aberrant structures, even in cells with as yet normal morphology (Extended Data Fig. 6a and compare Supplementary Videos 4 and 5). PG synthesis colocalized with mispositioned FtsZ (Extended Data Fig. 6b), consequently promoting disorderly cell wall synthesis and leading to morphological defects or cell lysis (Extended Data Fig. 6a and Supplementary Video 5). The remaining 71% of cells had an overrepresentation of abnormally large Z rings (diameter >1,000 nm) and an underrepresentation of Z rings with diameter <800 nm (55% in Δ*mapZ* compared with 75% in wild type) (Fig. 3f). The reduced occurrence of constricting Z rings and the decrease in cells harbouring FtsZ dumbbells (Fig. 3d and Extended Data Fig. 6c) suggest a premature closing of the septal Z ring in Δ*mapZ*. This is further supported by 3D-SIM imaging of 4′,6-diamidino-2-phenylindole (DAPI)-stained nucleoids, which revealed a very dense stretch of DNA trapped at the septum in 19% of Δ*mapZ* division figures (Fig. 3g and Supplementary Video 6). These chromosome pinching events most probably result from the previously inferred premature closing of the septum in an organism that lacks a nucleoid occlusion system[4]. Perhaps less surprisingly, cells with aberrant FtsZ structures also showed aberrant nucleoid shapes (Extended Data Fig. 6d and Supplementary Video 7). Therefore, not only is MapZ required for correct positioning of the Z ring but it is also involved in the regulation of constriction.

Co-immunoprecipitation revealed *in vivo* interaction between FtsZ and MapZ, which is mediated by the cytoplasmic domain of MapZ (Fig. 4a). The cytoplasmic domain of MapZ (MapZ_cyto) strongly interacts with FtsZ (affinity (K_D) = 8.76 nM), more precisely through its N-terminal peptide, which is predicted to be an α-helix (MapZ_{(1–41)}, from Met 1 to Gly 41) (K_D = 20.4 nM) (Extended Data Fig. 7a–d). Other parts of the intracellular domain (MapZ_{(42–98)} and MapZ_{(42–158)}) or the extracellular domain (MapZ_extra) showed no interactions with FtsZ (Extended Data Fig. 7e–g). *In vivo* deletion of the MapZ N-terminal peptide (*mapZΔ(1–41)*) did not impair MapZ septal localization (Extended Data Fig. 6e), but resulted in delocalization of FtsZ (Extended Data Fig. 6f, g), which subsequently leads to aberrant cell morphogenesis, asymmetric division or cell lysis (Extended Data Figs 1a, 6h and Supplementary Video 8). This was also observed in the *mapZΔcyto* strain (Extended Data Fig. 6i and
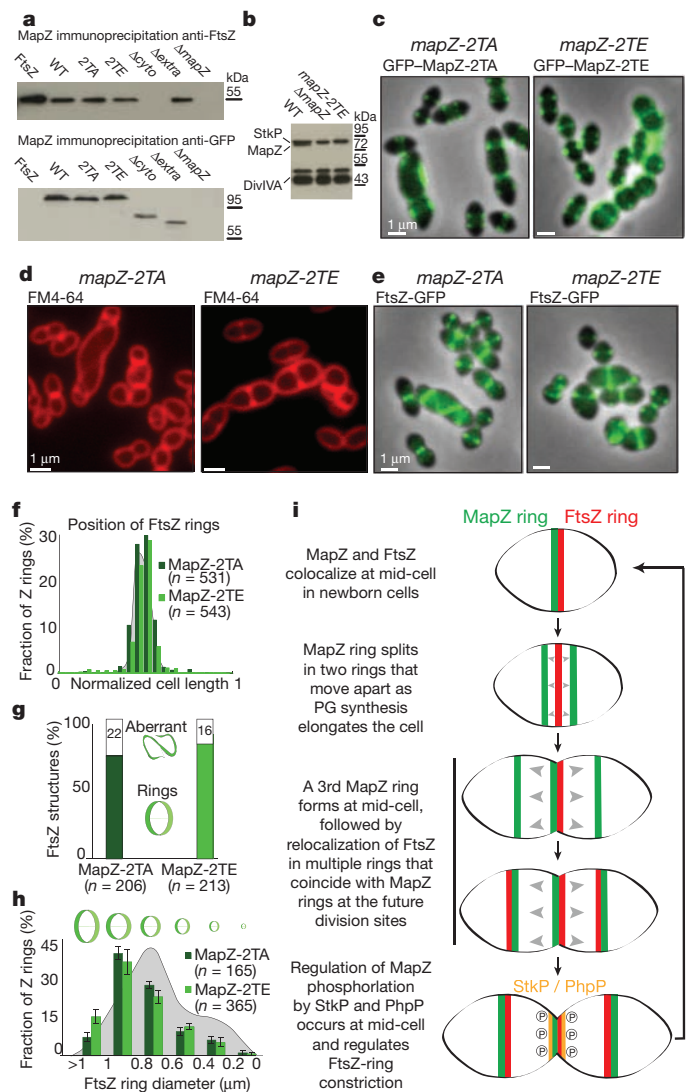
Figure 4 | Characterization of MapZ phosphorylation.
a, Immunoprecipitation of GFP–MapZ with FtsZ. Anti-FtsZ (top) or anti-GFP (bottom). WT, wild type. **b**, Western blot of cell lysates probed with anti-phosphothreonine antibodies shows phosphorylation signal for MapZ, DivIVA and StkP. **c**, Localization of GFP–MapZ-2TA and GFP–MapZ-2TE. **d**, *mapZ-2TA* and *mapZ-2TE* cells after membrane staining. **e**, Localization of FtsZ–GFP. **f**, Positioning of single FtsZ rings. **g**, Fraction of cells with FtsZ rings or aberrant structures (P value < 1.54×10^{-10} for MapZ-2TA and 1.14×10^{-10} for MapZ-2TE, two-tailed Student's *t*-test). **h**, Distribution of FtsZ-ring diameters (P value = 2.9×10^{-4} for MapZ-2TA and 9.9×10^{-3} for MapZ-2TE, two-tailed Student's *t*-test). **i**, Model of MapZ-mediated control of FtsZ positioning (sample of n cells analysed). Images are representative of experiments made in triplicate.

Supplementary Video 9). Therefore, the direct interaction of MapZ with FtsZ is strictly required for FtsZ positioning. Surprisingly, the conserved FtsZ C-terminal tail (Asp 408 to Arg 419), which promotes interaction with FtsZ regulators such as FtsA, EzrA, ZipA and SepF in various bacteria[17–20], is not required for interaction with MapZ (Extended Data Fig. 7h).

In agreement with a previous report[13], we confirmed MapZ phosphorylation by analysing the phosphorylation pattern of MapZ mutants (Fig. 4b). Mass spectrometry analysis of MapZ further showed that MapZ is phosphorylated on Thr 67 and Thr 78 (Extended Data Fig. 8a, b). We then constructed two mutants encoding either the phosphoablative form of MapZ (*mapZ-2TA*) or the phosphomimetic form (*mapZ-2TE*). *mapZ-2TA* and *mapZ-2TE* exhibited cell shape and viability defects (Fig. 4c, d, Extended Data Table 1 and Extended Data Fig. 1a), although

MapZ-2TA and MapZ-2TE retained septal localization (Fig. 4c) and the FtsZ rings were largely well positioned (Fig. 4e, f). Thus, FtsZ positioning by MapZ still occurs properly in *mapZ-2TA* and *mapZ-2TE*, consistent with the observation that phosphorylation does not affect interaction with FtsZ *in vitro* (Fig. 4a and Extended Data Fig. 7i). However, *mapZ-2TA* and *mapZ-2TE* showed aberrant FtsZ structures (Fig. 4g), altered Z-ring diameters (Fig. 4h) and a reduced number of FtsZ rings per cell (Fig. 3g). We conclude that both phosphorylated and dephosphorylated MapZ forms *in vivo*, and that most probably the balance between the two has a role in the control of FtsZ splitting, stability and constriction, but not in positioning.

This work uncovers a novel mechanism in which a single protein has the dual role of marking the cell division site and positioning the FtsZ ring (Fig. 4i). Our data are consistent with a model in which MapZ is anchored at the cell equator by its extracellular domain, which interacts with PG. It is possible that MapZ recognizes a PG structure specific to the mid-cell, such as the equatorial mark visible at the pneumococcal surface (Fig. 1b), reminiscent of the 'piecrust' previously reported in *Staphylococcus aureus*, *S. pneumoniae* and *Enterococcus faecalis*[6,21–23]. As PG synthesis forms the new cell halves, MapZ remains permanently associated with the equators, thus providing a simple mechanism to signal the site of division. MapZ intracellular domains on the inner side of the membrane act as a physical anchor, which positions the FtsZ ring at the division site. Subsequently, MapZ phosphorylation regulates cytokinesis, either through direct regulation of FtsZ or through regulation of other division factors. The fact that MapZ phosphorylation does not occur in the domain that interacts with FtsZ but in the neighbouring one (Extended Data Fig. 7b), and does not affect FtsZ polymerization or GTPase activity (Extended Data Fig. 8c, d), favours the idea of indirect regulation. MapZ cyclic phosphorylation/dephosphorylation most probably occurs when MapZ colocalizes with StkP at mid-cell (Extended Data Fig. 9a) and where the cytoplasmic phosphatase PhpP, which dephosphorylates MapZ (Extended Data Fig. 9b), is enriched[7]. We further hypothesize that MapZ-mediated recruitment of FtsZ is the event that initiates the assembly of the other division proteins at the septum, including GpsB, which has previously been shown[9] to be required for septal positioning of StkP and for the ability of StkP to phosphorylate its targets. Specifically, orchestrated phosphorylation of DivIVA and MapZ enables coordination between PG synthesis and control of the Z ring, respectively.

MapZ is conserved amongst *Streptococcaceae* and most other *Lactobacillales* (Extended Data Fig. 9c). These organisms lack homologues of known FtsZ regulatory systems[2–5]. Thus, the MapZ-mediated mechanism we have uncovered illustrates that pathways of cell division are far more diverse than previously thought in bacteria; they have adapted to the variety of bacterial lifestyles, cell shapes and developmental behaviours.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Adams, D. W. & Errington, J. Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nature Rev. Microbiol.* **7,** 642–653 (2009).
2. Bramkamp, M. & van Baarle, S. Division site selection in rod-shaped bacteria. *Curr. Opin. Microbiol.* **12,** 683–688 (2009).
3. Wu, L. J. & Errington, J. Nucleoid occlusion and bacterial cell division. *Nature Rev. Microbiol.* **10,** 8–12 (2012).
4. Pinho, M. G., Kjos, M. & Veening, J. W. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. *Nature Rev. Microbiol.* **11,** 601–614 (2013).
5. Monahan, L. G. & Harry, E. J. Identifying how bacterial cells find their middle: a new perspective. *Mol. Microbiol.* **87,** 231–234 (2013).
6. Monahan, L. G., Liew, A. T. F., Bottomley, A. L. & Harry, E. J. Division site positioning in bacteria: one size does not fit all. *Front. Microbiol.* **5,** 19 (2014).
7. Beilharz, K. *et al.* Control of cell division in *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proc. Natl Acad. Sci. USA* **109,** E905–E913 (2012).
8. Fleurie, A. *et al.* Mutational dissection of the S/T-kinase StkP reveals crucial roles in cell division of *Streptococcus pneumoniae*. *Mol. Microbiol.* **83,** 746–758 (2012).
9. Fleurie, A. *et al.* Interplay of the serine/threonine-kinase StkP and the paralogs DivIVA and GpsB in pneumococcal cell elongation and division. *PLoS Genet.* **10,** e1004275 (2014).
10. Hanks, S. K., Quinn, A. M. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241,** 42–52 (1988).
11. Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. & Manning, G. Structural and functional diversity of the microbial kinome. *PLoS Biol.* **5,** e17 (2007).
12. Pereira, S. F., Goss, L. & Dworkin, J. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol. Mol. Biol. Rev.* **75,** 192–212 (2011).
13. Nováková, L. *et al.* Identification of multiple substrates of the StkP Ser/Thr protein kinase in *Streptococcus pneumoniae*. *J. Bacteriol.* **192,** 3629–3638 (2010).
14. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37,** W465–W468 (2009).
15. Kuru, E. *et al. In situ* probing of newly synthesized peptidoglycan in live bacteria with fluorescent D-amino acids. *Angew. Chem. Int. Edn Engl.* **51,** 12519–12523 (2012).
16. Daniel, R. A. & Errington, J. Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell. *Cell* **113,** 767–776 (2003).
17. Ma, X. & Margolin, W. Genetic and functional analyses of the conserved C-terminal core domain of *Escherichia coli* FtsZ. *J. Bacteriol.* **181,** 7531–7544 (1999).
18. Singh, J. K., Makde, R. D., Kumar, V. & Panda, D. A membrane protein, EzrA, regulates assembly dynamics of FtsZ by interacting with the C-terminal tail of FtsZ. *Biochemistry* **46,** 11013–11022 (2007).
19. Haney, S. A. *et al.* Genetic analysis of the *Escherichia coli* FtsZ.ZipA interaction in the yeast two-hybrid system. Characterization of FtsZ residues essential for the interactions with ZipA and with FtsA. *J. Biol. Chem.* **276,** 11980–11987 (2001).
20. Król, E. *et al. Bacillus subtilis* SepF binds to the C-terminus of FtsZ. *PLoS ONE* **7,** e43293 (2012).
21. Wheeler, R., Mesnage, S., Boneca, I. G., Hobbs, J. K. & Foster, S. J. Super-resolution microscopy reveals cell wall dynamics and peptidoglycan architecture in ovococcal bacteria. *Mol. Microbiol.* **82,** 1096–1109 (2011).
22. Tomasz, A., Jamieson, J. D. & Ottolenghi, E. The fine structure of *Diplococcus pneumoniae*. *J. Cell Biol.* **22,** 453–467 (1964).
23. Higgins, M. L. & Shockman, G. D. Model for cell wall growth of *Streptococcus faecalis*. *J. Bacteriol.* **101,** 643–648 (1970).

Supplementary Information is available in the online version of the paper.

Author Contributions C.G. and C.L. wrote the paper. A.F., C.L., S.M., C.Z., C.Cl., J.-P.L., M.F.-W. and C.G. designed and performed the experiments. C.G., C.L., A.F., S.M., J.-P.L., C.Co., M.F.-W. and B.M. analysed and interpreted the data. E.K., Y.V.B. and M.S.VN. provided new reagents. E.K., Y.B. and D.S. read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.G. (c.grangeasse@ibcp.fr) or C.L. (christian.lesterlin@bioch.ox.ac.uk).

METHODS

Strains, plasmids, primers and growth conditions. *S. pneumoniae* isogenic strains were constructed by transformation in R800 (ref. 8). Standard procedures for chromosomal transformation and growth were used[9,24]. For growth experiments, *S. pneumoniae* strains were cultivated at 37 °C in Todd–Hewitt yeast (THY) broth (Difco). For induction of P_{Zn}, $ZnCl_2$ was added at the concentration of 0.2 mM. For construction of *S. pneumoniae* mutants, transformation was performed as described previously using precompetent cells treated at 37 °C with synthetic competence stimulating peptide 1 (CSP1) to induce competence[8,25]. For viability assays, several samples of exponentially growing cells were taken every 30 min, diluted appropriately and plated onto THY agar supplemented with horse blood. After overnight incubation, colony-forming units (c.f.u.) were counted and the percentage of viability of mutant strains was expressed relative to the wild-type strain. These experiments were biologically made in triplicates. The *Escherichia coli* XL1-Blue strain was used as a host for cloning. *E. coli* BL21(DE3) strain was used as host for overexpression. Strains used in this study are listed in Supplementary Table 1.

Construction of plasmids. DNA fragments coding for full-length MapZ, the extracellular domain, the cytoplasmic domain or peptides of MapZ, the kinase domain of StkP, PhpP and for FtsZ$_{(1-407)}$, were obtained by PCR using chromosomal DNA from *S. pneumoniae* R800 strain as a template and primer pairs 47/48, 39/40, 34/35, 34/35, 34/36, 37/38, 37/35, 45/46, 41/42 and 43/44, respectively (Supplementary Table 2). DNA fragment coding for the cytoplasmic domain of MapZ with T67-78E mutations was obtained using primer pair 34/35 and the *mapZ-2TE* strain as a template (Supplementary Table 1). *mapZ$_{cyto}$*, *mapZ$_{cyto}$-2TE*, *mapZ$_{(1-41)}$*, *mapZ$_{(42-98)}$*, *mapZ$_{(42-158)}$*, *mapZ$_{extra}$* and *phpP* were cloned between the NdeI and PstI cloning sites of the pT7-7 plasmid[26]. *ftsZ$_{(1-407)}$* was cloned between the NdeI and BamHI cloning sites of the pETPhos plasmid[27]. *mapZ* was cloned between the AgeI and NotI cloning sites of the pCM38 plasmid (gift from C. Morlot, see Supplementary Table 1). *stkP$_{KD}$* was cloned between the BamHI and HindIII cloning sites of the pQE30 plasmid. The nucleotide sequences of all DNA fragments were checked to ensure error-free amplification.

Allelic replacement mutagenesis. To construct pneumococcus mutants (gene deletions, *gfp/rfp* fusions or site-directed mutagenesis), we used a two-step procedure, based on a bicistronic *kan-rpsL* cassette called Janus[24]. The genes encoding RFP and GFP were from refs 7 and 28, respectively. The Janus procedure allows the replacement of a gene by a cassette and subsequent deletion or substitution of the cassette by a mutated allelic form at the gene chromosomal locus. This procedure avoids polar effects and allows a physiological level of expression of GFP and RFP fusions and mutated proteins. Briefly, the Janus cassette is either used to replace the gene of interest or inserted at either its 5′ or 3′ end. Both options confer resistance to kanamycin and dominant streptomycin sensitivity in the wild-type streptomycin-resistant R800 *rpsL1* strain (Kan^R-Str^S). Then, any DNA fragments flanked on each end by sequences homologous to the upstream and downstream regions of the gene of interest could be used to transform Kan^R-Str^S strains in order to obtain the expected nonpolar markerless mutant strains Kan^S-Str^R. Once obtained, these markerless transformants were re-streaked to single colonies and correct integration at the chromosomal locus was verified by PCR. Full description of primers used for the construction of strains (Supplementary Table 1) is provided in Supplementary Table 2.

Protein purification. Recombinant plasmids overproducing MapZ$_{cyto}$, MapZ$_{cyto}$-T67-78E, MapZ$_{(1-41)}$, MapZ$_{(42-98)}$, MapZ$_{(42-158)}$, MapZ$_{extra}$, FtsZ$_{(1-407)}$, FtsZ, StkP$_{KD}$ and PhpP were transformed into the BL21(DE3) *E. coli* strain. Overexpression and purification of StkP$_{KD}$, and FtsZ and FtsZ$_{(1-407)}$, were performed as previously described in refs 8 and 29, respectively. MapZ wild-type or mutated domains as well as PhpP were purified using the same procedure than StkP$_{KD}$. To purify MapZ from *S. pneumoniae* cells, we used the strain in which *mapZ* is fused to a DNA fragment encoding 6 histidines (Supplementary Table 1) and the procedure was carried out as previously described[9]. We checked that cells grew as the wild-type cells and displayed proper cell shape.

Peptidoglycan labelling with fluorescent D-amino acids. The procedure used was adapted from[15]. Briefly, exponentially growing *gfp-mapZ* or *ftsZ- gfp* strains (OD$_{550 \, nm}$ = 0.1) were incubated for 1 min at 37 °C in THY with 500 μM of TDL (a fluorescent carboxytetramethylrhodamine derivative of D-alanine). Cells were then washed three times with 1 ml PBS pH 7.4 at room temperature, incubated again for 1 min at 37 °C with 500 μM of HADA (a fluorescent hydroxy coumarin derivative of D-alanine) of OD$_{550 \, nm}$ = 0.1 and washed three times with PBS. For localization of FtsZ together with PG synthesis, *ΔmapZ ftsZ-gfp* cells were grown up to OD$_{550 \, nm}$ = 0.1 in THY and labelled for 3 min with 500 μM of TDL, and finally washed three times with PBS. 0.7 μl of each mixture was then placed on slides and observed under the microscope. These experiments were biologically made in triplicates.

Microscopy techniques. Microscopy was performed on exponentially growing cells ($A_{550 \, nm}$ = 0.1). TEM, SEM, fluorescence and immunofluorescence microscopy

were carried out as previously described[8]. Slides were visualized with a Zeiss Axio-Observer Z1 microscope fitted with an Orca-R2 C10600 charge-coupled device (CCD) camera (Hamamatsu) with a ×100 NA 1.46 objective. Images were collected with AxioVision (Carl Zeiss). For TEM, cells were examined with a Philips CM120 transmission electron microscope equipped with a Gatan Orius SC200 CCD camera. For SEM, cells were observed with a Quanta 250 FEG (FEI) scanning electron microscope. Time-lapse microscopy was performed as described[30] using an automated inverted epifluorescence microscope Nikon Ti-E/B equipped with the perfect focus system (PFS, Nikon) and a phase contrast objective (CFI Plan Fluor DLL ×100 oil NA 1.3), a Semrock filter set for GFP (Ex: 482BP35; DM: 506; Em: 536BP40), a Nikon Intensilight 130W High-Pressure Mercury Lamp, a monochrome OrcaR2 digital CCD camera (Hamamatsu) and an ImagEM-1K EMCCD camera (Hamamatsu). The microscope is equipped with a chamber thermostated at 30 °C. Images were captured every 5 min and processed using Nis-Elements AR software (Nikon). All fluorescence images were acquired with a minimal exposure time to minimize bleaching and phototoxicity effects. GFP fluorescence images were false coloured green and overlaid on phase contrast images. Super-resolution 3D-SIM imaging was carried out as previously described[31], on a DeltaVision OMX V3 (Applied Precision/GE Healthcare) equipped with a Blaze SIM module, a ×60/1.42 oil UPlanSApo objective (Olympus), 405 nm and 488 nm diode lasers and three sCMOS cameras (PCO). Each 3D-SIM stacks is composed of 225 images (512 × 512 pixels) consisting of 12 z-sections (125 nm z-distance), with 15 images per z-section with the striped illumination pattern[32,33] rotated to the three angles (−60°, 0°, +60°) and shifted in five phase steps. Acquisition settings were as follows: for FtsZ–GFP, 3 ms exposure with 488 nm laser (attenuated to 100% transmission); GFP–MapZ, 7 ms exposure with 488 nm laser (attenuated to 100% transmission); DAPI, 20–30 ms exposure with 405 nm laser (100% transmission). The 3D-SIM raw data was reconstructed with SoftWoRx 6.0 (Applied Precision) using a Wiener filter setting of 0.002 and channel specifically measured optical transfer functions resulting in a lateral (x–y) resolution of 100–130 nm (wavelength dependent) and an axial (z) resolution of ~300. These experiments were technically made in triplicates.

Microscopy image analysis. Snapshot analysis was performed using ImageJ (http://rsb.info.nih.gov/ij/) and the MicrobeTracker suite[34] extended by custom MATLAB routines to generate cell length and width distribution histograms, fluorescent intensity linescans, focus positioning dotplots and histograms, cumulative distributions of cells with 1, 2 and 3 rings, and plots of inter-ring distance (IRD) as a function of cell length. Ring diameter measurements were performed using SoftWoRx 6.0 (Applied Precision). We performed Student's *t*-tests for statistical analysis of our data using StatPlus plug-in for Excel-Mac (by AnalystSoft), which provided the two-tailed distribution *P* values given in the figure legends (with a critical value of 0.05). For the cell length distribution analysis presented in Extended Data Fig. 1, a non-parametric statistical analysis was performed, as detailed previously[35,36], to take into account the non-normal distribution of cell sizes in the mutants strains analysed.

Co-immunoprecipitation of FtsZ and GFP–MapZ with anti-GFP antibodies. Cultures of *S. pneumoniae* cells were grown at 37 °C in THY medium until OD$_{550 \, nm}$ = 0.4. Cell pellets were incubated at 30 °C first for 30 min in buffer A (0.1 mM Tris-HCl, 2 mM MgCl$_2$, 1 M sucrose, 1:100 Protease Inhibitory Cocktail, 1 mg ml^{-1} of DNase I and RNase A) and then in buffer B (0.1 mM Tris-HCl, 1 mM EDTA, 1% (v/v) Triton X-100, 1:100 Protease Inhibitory Cocktail, 1 mg ml^{-1} of DNase I and RNase A) at room temperature for 15 min. After centrifugation, the supernatant was incubated with the GFP-Trap resin suspension according to the manufacturer's instructions (Chromotech). Protein-bound GFP-Trap resins were eluted with Laemmli buffer at 95 °C for 10 min and analysed by SDS–PAGE. These experiments were biologically and technically made in triplicates.

FtsZ polymerization and GTPase assays. FtsZ polymerization assays were performed as described previously[29]. Briefly, mixtures of 3 μM of FtsZ and 6 μM of MapZ, MapZ-2TA or MapZ-2TE were incubated for 15 min at 25 °C in a buffer containing 50 mM HEPES/NaOH, pH 7.2, 50 mM KCl, 10 mM MgCl$_2$ and 1 mM β-mercaptoethanol. Identical reaction conditions were ensured by compensating varying amounts of proteins with the storage buffer. The solutions were subsequently centrifuged for 15 min at 250,000g and 25 °C in a Beckman 50.4 Ti rotor using a Beckmann LE80K ultracentrifuge. After immediate withdrawal of the supernatants, pelleted proteins were dissolved in 150 μl SDS sample buffer and incubated for 10 min at 96 °C. Fifteen microlitres of each sample were then subjected to electrophoresis in a 10% SDS–PAGE gel. For visualization, gels were stained with Brilliant Blue R 250 and scanned. These experiments were technically made in triplicates. GTPase assays were performed following the previously described procedure[37]. The reaction was performed in a buffer containing 50 mM HEPES/NaOH pH 7.5 and 300 mM KCl. Master mixes contained 24 μM of FtsZ and when needed 48 μM of MapZ$_{cyto}$ or MapZ$_{cyto}$-2TE.

MapZ *in vitro* phosphorylation and dephosphorylation. *In vitro* phosphorylation of MapZ$_{cyto}$ by StkP$_{KD}$ was carried out by incubating the reaction mixture (200 μl) containing 50 μg of MapZ$_{cyto}$, 1 μg StkP$_{KD}$ and 25 mM Tris-HCl, pH 7.0, 1 mM

dithiothreitol (DTT), 5 mM $MgCl_2$, 1 mM EDTA and 10 μM ATP with 5 μCi [γ-^{32}P]ATP (specific activity 3,000 Ci mmol^{-1}) for 15 min at 37 °C. Twenty microlitres were sampled and mixed with SDS–PAGE loading buffer and heated for 5 min at 100 °C. The remaining mixture was further incubated in the presence of 0.2 μg of purified PhpP and 20 μl aliquots were withdrawn at 30 s, 1, 2, 5, 10 min and mixed with SDS–PAGE loading buffer and heated. After SDS–PAGE analysis, gels were soaked in 16% trichloroacetic acid for 10 min at 90 °C, stained with Coomassie blue, dried and MapZ$_{cyto}$ dephosphorylation was visualized by autoradiography using X-ray films (Kodak BIOMAX-MS). These experiments were technically made in triplicates.

MapZ cell wall binding. Pneumococcal cell wall preparation as well as the procedure used to analyse MapZ binding to the cell wall was described previously[38]. Briefly, 2 μg of purified MapZ$_{extra}$ was incubated with purified *S. pneumoniae* cell wall (5 mg) in 100 μl of a buffer containing 50 mM Tris pH 8.0 and 100 mM NaCl for 16 h at 4 °C. After centrifugation (5 min at 5,000*g*), the supernatant was removed (unbound fraction) and the cell wall pellet was washed three times with PBS (wash fraction) and resuspended in 50 μl SDS–PAGE loading buffer. After incubation at 100 °C for 10 min, the supernatant, corresponding to the cell wall bound to MapZ (bound fraction), was recovered from the cell wall pellet by centrifugation (5 min at 5,000*g*). The different fractions were analysed by SDS–PAGE and western immunoblotting. These experiments were technically made in triplicates.

Immunoblot analysis. *In vivo* phosphorylated proteins in crude extracts of *S. pneumoniae* strains were immunodetected using an anti-phosphothreonine polyclonal antibody (Cell Signaling) at 1/2,000 as described previously[8]. For FtsZ, immunodetection was performed using a specific rabbit polyclonal antibody[39] used at 1/10,000. Detection of GFP fusions was performed using a rabbit anti-GFP antibody (AMS Biotechnology). Detection of MapZ$_{extra}$ in cell wall binding assays was performed using a mouse anti-6×His antibody (Sigma). A goat anti-rabbit secondary antibody horseradish peroxidase (HRP) conjugate (Biorad) was used at 1/5,000 to reveal the immunoblots, except for the cell wall binding assay, in which goat anti-mouse secondary antibody HRP conjugate (Biorad) was used at 1/5,000. These experiments were biologically and technically made in triplicates.

Nano-LC-MS/MS analysis of purified MapZ. Purified MapZ was in gel digested using trypsin as described elsewhere[40]. Peptide mixture was either analysed directly by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) after being desalted using C18 StageTips[41] or was subjected to phosphopeptide enrichment by titanium dioxide chromatography as described previously[42] with the following modifications: phosphopeptide elution from the beads was performed three times with 100 ml 40% ammonia hydroxide solution in 60% acetonitrile at a pH of >10.5. Analysis of peptides and phosphopeptides was done on a Proxeon Easy-LC system (Proxeon Biosystems) either coupled to an LTQ-Orbitrap-Elite or to an LTQ-Orbitrap-XL mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source (Proxeon Biosystems) as described previously[43]. The mass spectrometer was operated in the positive ion mode with the following acquisition cycle: one initial full scan in the Orbitrap analyser (MS) was followed by fragmentation through rapid collision induced dissociation (CID) of the 20 most intense multiply charged precursor ions in the linear ion trap analyser (LTQ Elite), or the five most intense precursor ions (LTQ XL). Here multi-stage activation (MSA) was applied in all MS/MS events when a neutral loss event was detected on the precursor ions depending on their charge state: singly (−97.97 Th), doubly (−48.99 Th) and triply (−32.66 Th). Mass spectra were analysed using the software suite MaxQuant, v.1.0.14.3 (ref. 44). The data were searched against a target-decoy *S. pneumoniae* database including the His-tagged sequence of MapZ (35,203 entries) and 262 commonly observed protein contaminants. Trypsin was set as protease and two missed cleavage sites were allowed. Acetylation at the N terminus, oxidation of methionine and phosphorylation on serine, threonine and tyrosine were set as variable modifications. Carbamidomethylation of cysteine was set as fixed modification. Initial precursor mass tolerance was set to 7 p.p.m. at the precursor ion level and 0.5 Da at the fragment ion level. Phosphorylation events with a localization probability of at least 0.75 were considered to be assigned to a specific residue. Spectra of modified peptides were manually validated.

Surface plasmon resonance. Real-time binding experiments were performed on a BIAcore T100 biosensor system (GE Healthcare). FtsZ and FtsZ$_{(1-407)}$ were covalently coupled through its amino groups to the surface of a CM5 sensorchip according to the manufacturer's instructions. Increasing concentrations (0.01, 0.02, 0.05, 0.1, 0.2 and 0.5 μM from bottom to top) of MapZ$_{cyto}$, MapZ(1–41), MapZ(42–98), MapZ(42–158), MapZ$_{extra}$ and MapZ$_{cyto}$-2TE were injected over the surface of the sensorchip at a flow rate of 30 μl min^{-1} in 10 mM HEPES pH 7.4, 150 mM NaCl, 0,005% surfactant P20. For all experiments, non-specific binding to the surface of the sensorchip was subtracted by injection of the analytes over a mocked derivatized sensorchip. The resulting sensorgrams were analysed using BIAevaluation software (GE Healthcare) according to a 1:1 model of interaction to determine the kinetic constants. The goodness of the fit was assessed by inspecting the χ^2 values

and the random distribution of the residuals. These experiments were technically made in triplicates.

MapZ sequence analysis and search of MapZ in bacterial genomes. MapZ topology was predicted using TOPSCONS (http://topcons.net)[14] and secondary structure prediction of MapZ$_{cyto}$ was computed using the Network Protein Sequence Analysis (NPSA) (http://npsa-pbil.ibcp.fr) using DSC, PHD and SOPMA methods[45]. MapZ sequences were extracted from UniProtKB[46] complete bacteria genomes by means of GGsearch v.36.3.5c[47] with UniProtKB/Swiss-Prot:Q8DR55 as the query sequence. A multiple sequence alignment was computed with MUSCLE v.3.8.31 (ref. 48) with the 66 extracted sequences. Then, a profile HMM was built with the hmmbuild program of the HMMER 3.0 package[49]. The predicted proteins of the 6,305 bacterial genomes from Ensemble Genomes release[50] were searched with the hmmsearch program and the built profile. Subject sequences were extracted from matches found thanks to an in-house Java program if they observed the following conditions: E value $\leq 1 \times 10^{-24}$ and length between 350 and 650 amino acids.

24. Sung, C. K., Li, H., Claverys, J. P. & Morrison, D. A. An *rpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **67,** 5190–5196 (2001).
25. Martin, B., Prudhomme, M., Alloing, G., Granadel, C. & Claverys, J. P. Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol. Microbiol.* **38,** 867–878 (2000).
26. Cortay, J. C. *et al. In vitro* asymmetric binding of the pleiotropic regulatory protein, FruR, to the ace operator controlling glyoxylate shunt enzyme synthesis. *J. Biol. Chem.* **269,** 14885–14891 (1994).
27. Canova, M. J., Kremer, L. & Molle, V. pETPhos: a customized expression vector designed for further characterization of Ser/Thr/Tyr protein kinases and their substrates. *Plasmid* **60,** 149–153 (2008).
28. Martin, B. *et al.* Expression and maintenance of ComD–ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol. Microbiol.* **75,** 1513–1528 (2010).
29. Thanbichler, M. & Shapiro, L. MipZ, a spatial regulator coordinating chromosome segregation with cell division in *Caulobacter*. *Cell* **126,** 147–162 (2006).
30. de Jong, I. G., Beilharz, K., Kuipers, O. P. & Veening, J. W. Live cell imaging of *Bacillus subtilis* and *Streptococcus pneumoniae* using automated time-lapse microscopy. *J. Vis. Exp.* **53,** 3145 (2011).
31. Lesterlin, C., Ball, G., Schermelleh, L. & Sherratt, D. J. RecA bundles mediate homology pairing between distant sisters during DNA break repair. *Nature* **506,** 249–253 (2014).
32. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320,** 1332–1336 (2008).
33. Gustafsson, M. G. *et al.* Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94,** 4957–4970 (2008).
34. Sliusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80,** 612–627 (2011).
35. Krzywinski, M. & Altman, N. Points of significance: significance, *P* values and *t*-tests. *Nature Methods* **10,** 1041–1042 (2013).
36. Krzywinski, M. & Altman, N. Points of significance: importance of being uncertain. *Nature Methods* **10,** 809–810 (2013).
37. Król, E. & Scheffers, D. J. FtsZ polymerization assays: simple protocols and considerations. *J. Vis. Exp.* **81,** e50844 (2013).
38. Morlot, C. *et al.* Interaction of penicillin-binding protein 2x and Ser/Thr protein kinase StkP, two key players in *Streptococcus pneumoniae* R6 morphogenesis. *Mol. Microbiol.* **90,** 88–102 (2013).
39. Morlot, C., Zapun, A., Dideberg, O. & Vernet, T. Growth and division of *Streptococcus pneumoniae*: localization of the high molecular weight penicillin-binding proteins during the cell cycle. *Mol. Microbiol.* **50,** 845–855 (2003).
40. Borchert, N. *et al.* Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* **20,** 837–846 (2010).
41. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2,** 1896–1906 (2007).
42. Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4,** 2010–2021 (2005).
43. Franz-Wachtel, M. *et al.* Global detection of protein kinase D-dependent phosphorylation events in nocodazole-treated human cells. *Mol. Cell. Proteomics* **11,** 160–170 (2012).
44. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26,** 1367–1372 (2008).
45. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25,** 147–150 (2000).
46. UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41,** D43–D47 (2013).
47. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85,** 2444–2448 (1988).
48. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5,** 113 (2004).
49. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **7,** e1002195 (2011).
50. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41,** D48–D55 (2013).
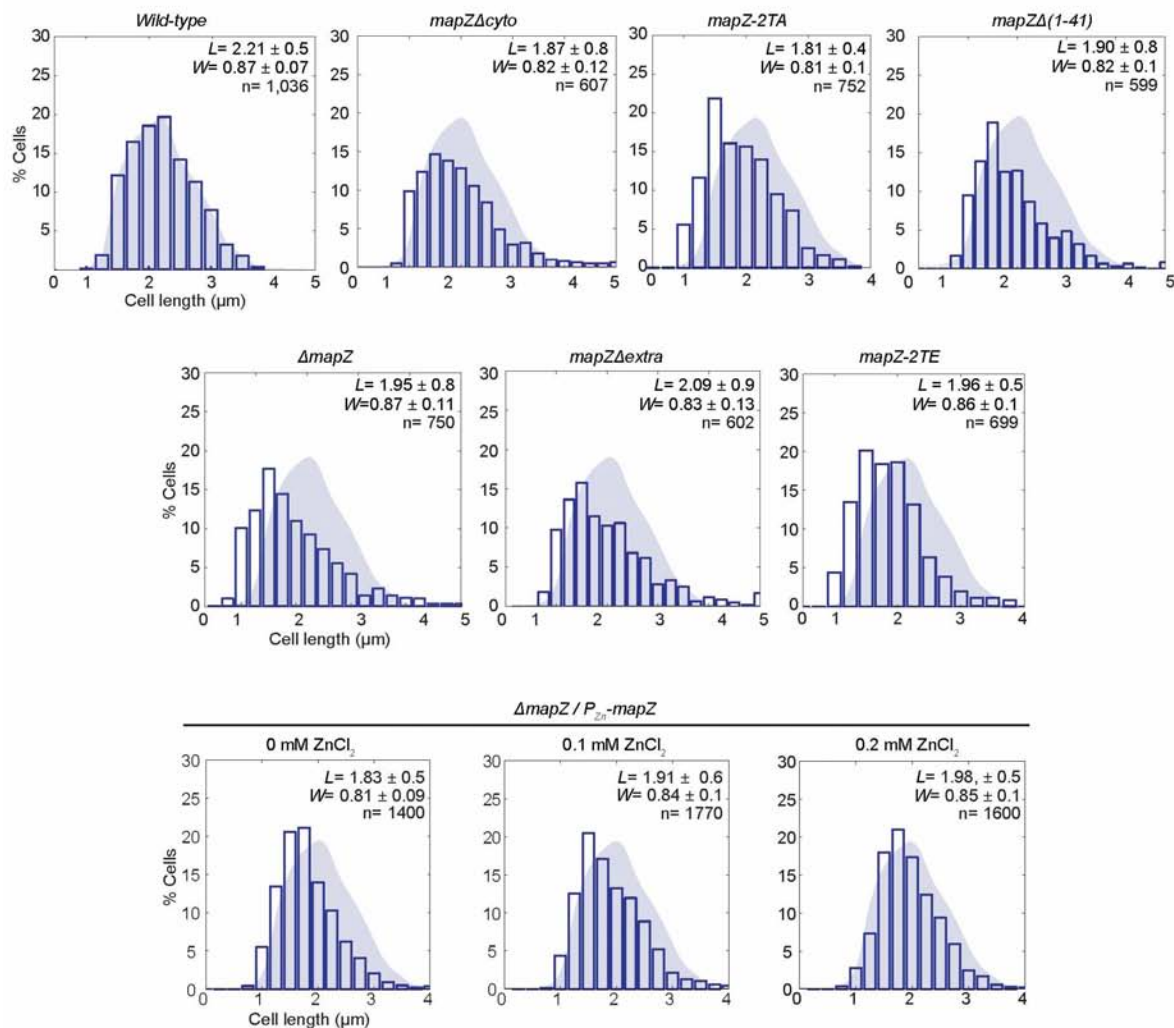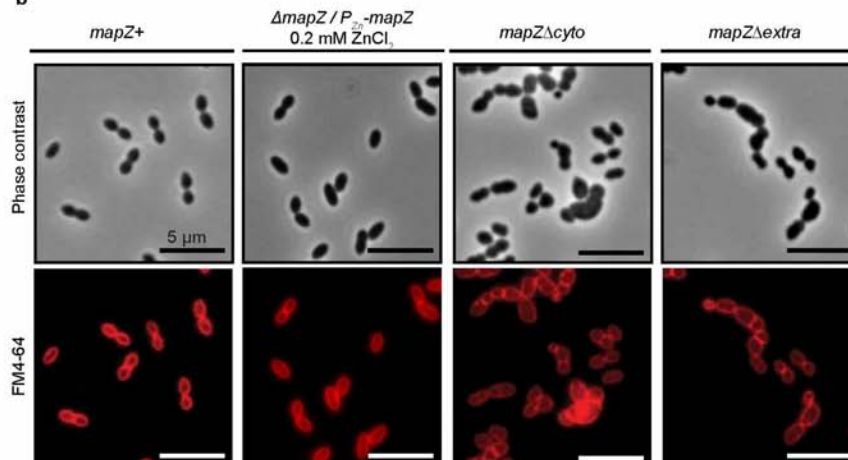
Extended Data Figure 1 | Cell-shape analysis of wild-type and *mapZ* mutant strains. a, Cell length distribution of wild-type, Δ*mapZ*, *mapZ-2TA*, *mapZ-2TE*, *mapZΔcyto*, *mapZΔextra* and *mapZΔ(1–41)* strains, as well as for Δ*mapZ*/P$_{Zn}$-*mapZ* in the presence of 0, 0.1 or 0.2 mM of ZnCl$_2$ inducer. Average cell length (*L*) and width (*W*) are given with standard deviations for a total of *n* cells analysed from three independent experiments.

P value $< 1.59 \times 10^{-2}$, two-tailed *t* distribution determined using a non-parametric statistical test for a critical value of 0.05. **b**, Phase contrast microscopy and FM4-64 membrane staining imaging of *mapZ*⁺ cells (*mapZ* is restored at the chromosomal locus in Δ*mapZ*), Δ*mapZ*/P$_{Zn}$-*mapZ* (Δ*mapZ* cells complemented ectopically with P$_{Zn}$-*mapZ*), *mapZΔcyto* and *mapZΔextra* cells. Images are representative of experiments made in triplicate.

a

b
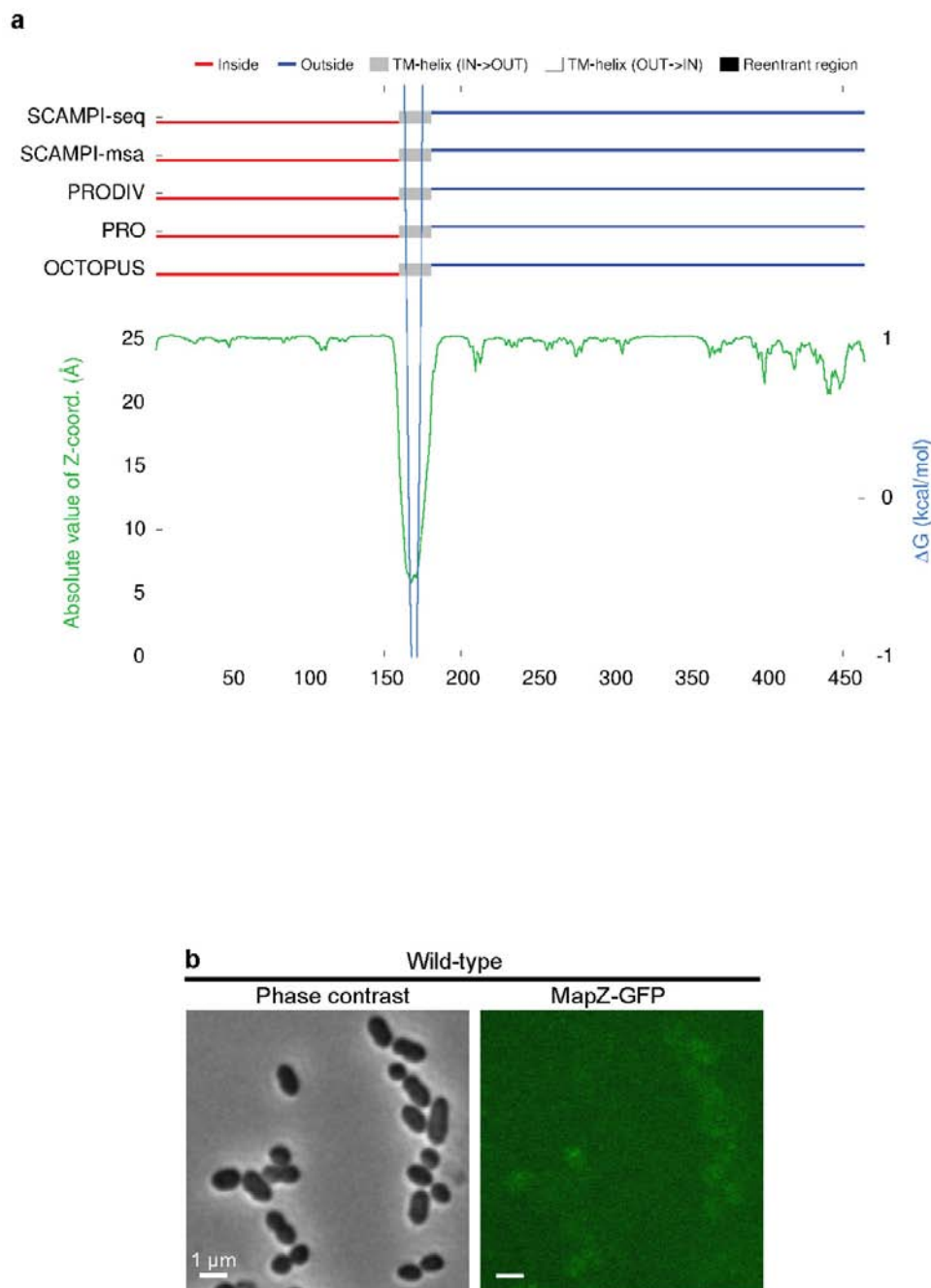
Extended Data Figure 2 | Prediction of MapZ topology. a, MapZ topology (that is, specification of the membrane spanning segments and their in/out orientation relative to the membrane) was predicted by five different topology algorithms (SCAMPI-seq, SACAMPI-msa, PRODIV-TMHMM, PRO-TMHMM and OCTOPUS) using TOPCONS (http://topcons.net). ZPRED (green line) predicts the distance to the membrane centre of each amino acid and ΔG scale (light blue) shows the predicted free energy of membrane insertion for a window of 21 amino acids centred around each position in the sequence. The transmembrane span is indicated in grey. Predictions of cytoplasmic and extracellular localizations are shown in red and dark blue, respectively. **b**, Wide-field microscopy images of cells producing the C-terminal fusion of MapZ with GFP. GFP fluorescence (right panel) and phase contrast images (left panel). Images are representative of experiments made in triplicate.
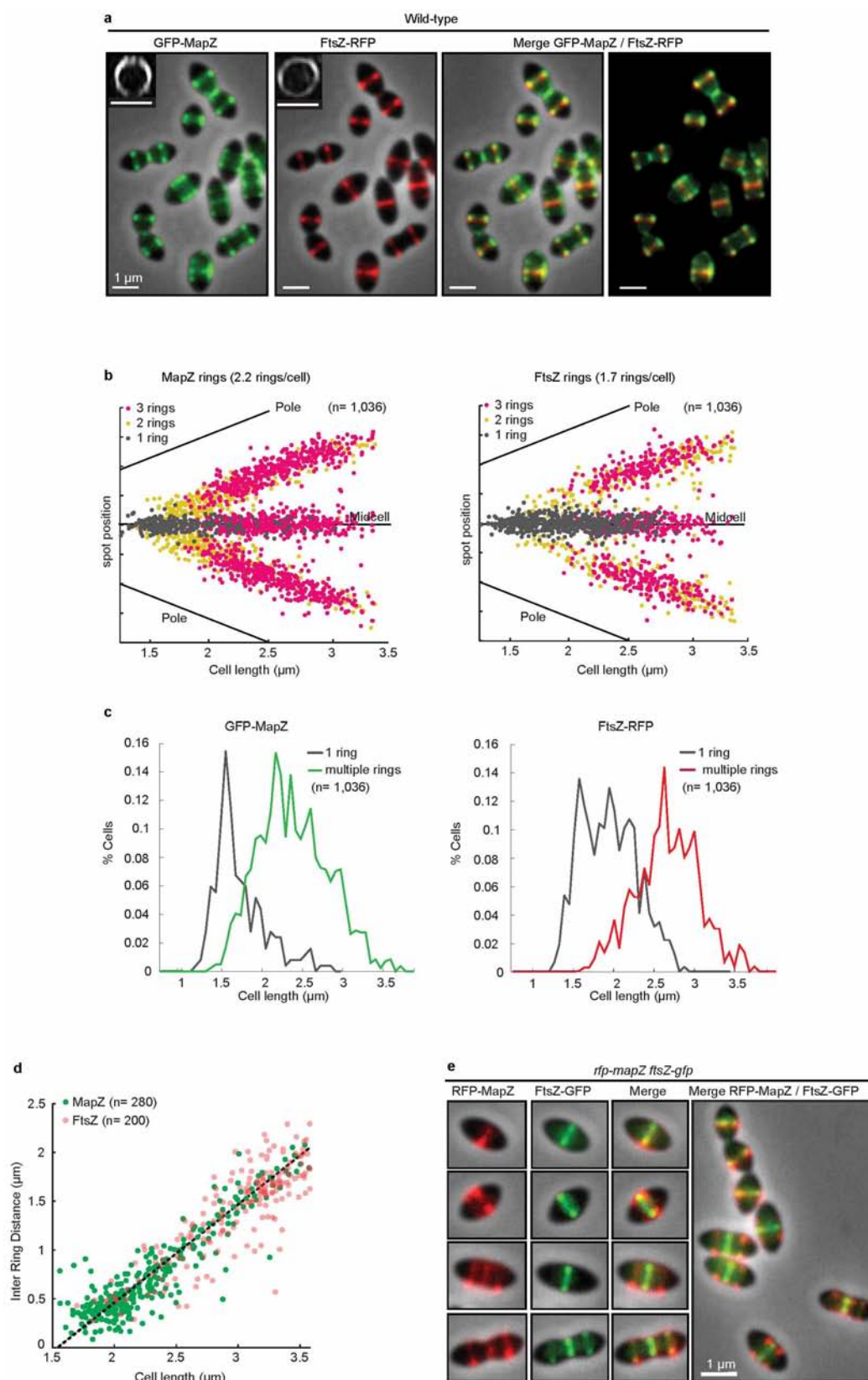
Extended Data Figure 3 | Localization of MapZ and FtsZ in wild-type cells. **a,** Microscopy images of GFP–MapZ and FtsZ–RFP in wild-type cells. Insert images show 3D-SIM orthogonal views of MapZ and FtsZ rings. **b,** Localization dotplots of MapZ-ring and FtsZ-ring positions along the cell length in wild-type cells. **c,** Ratio of cells with single or multiple MapZ rings and FtsZ rings as a function of cell length. **b, c,** Data are derived from analysis of 1,036 cells (*n* indicates the number of cells analysed in each panel). **d,** Distance between MapZ outer rings compared with distance between FtsZ outer rings as a function of cell length. **e,** Same as Fig. 2b but after swapping the GFP and RFP fluorescent protein labels. Indicative images showing MapZ, FtsZ, or both MapZ and FtsZ signals are shown for *rfp-mapZ ftsZ-gfp* cells at four different cell cycle stages. A wide-field view is also shown. Images are representative of experiments made in triplicate.
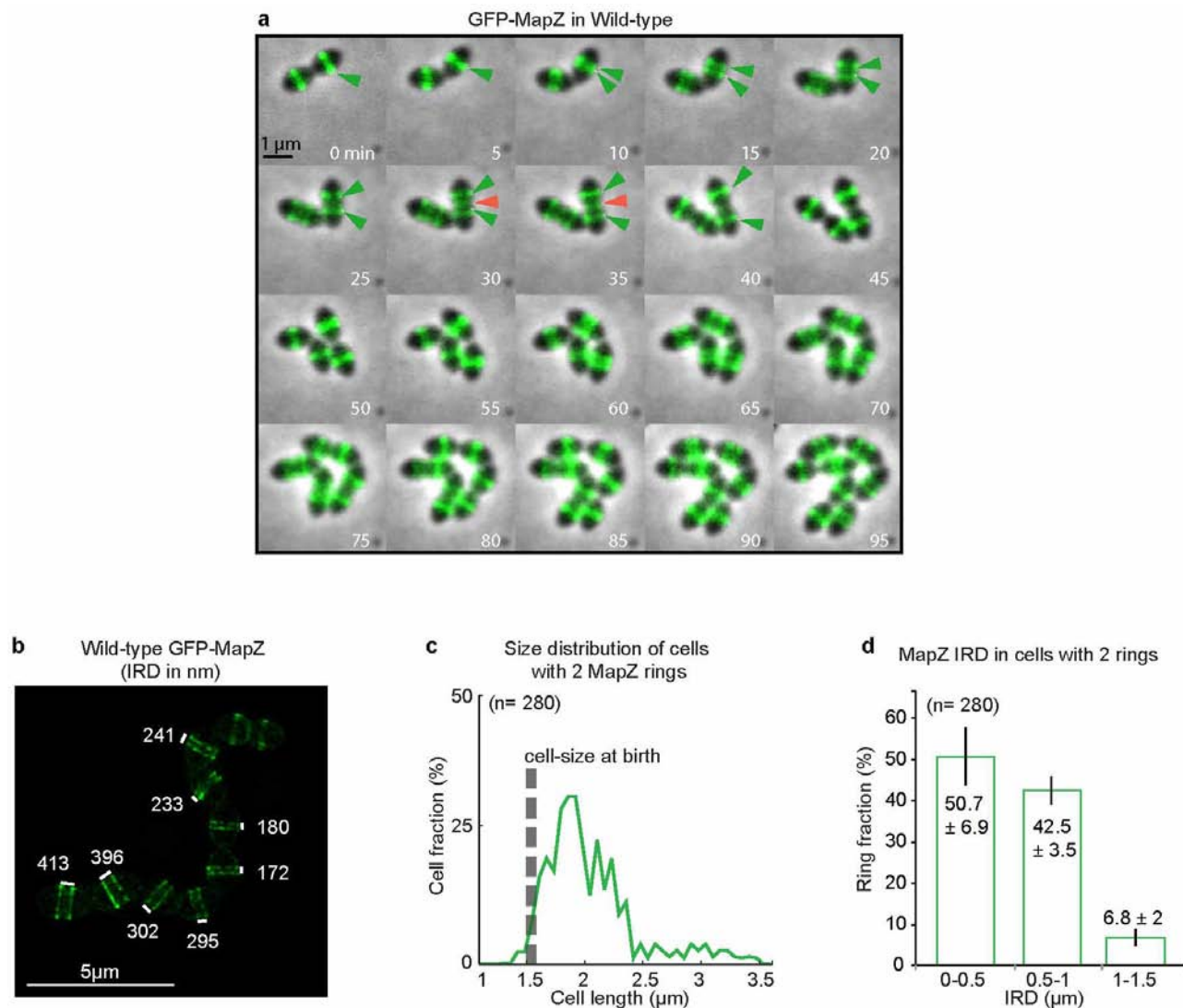
Extended Data Figure 4 | Early splitting of MapZ rings during elongation of wild-type cells. **a**, Time-lapse images of GFP–MapZ dynamics during cell growth and division showing progressive separation of the outer rings (green arrow) and appearance of a 3rd mid-cell ring (red arrow) (similar to Supplementary Video 1). Time is given in minutes. **b**, 3D-SIM images showing the very early stages of MapZ separation in the first stages of cell elongation. The numbers correspond to the inter-ring distance (IRD) in nm. **c**, Cell size distribution of cells with two MapZ rings reveals splitting of MapZ in the early stages of cells elongation. **d**, Distribution of IRD in cells with two MapZ rings (error bars show s.d. from three experiments). **c**, **d**, Data are derived from analysis of 280 cells (n indicates the number of cells analysed in each panel). Images are representative of experiments made in triplicate.
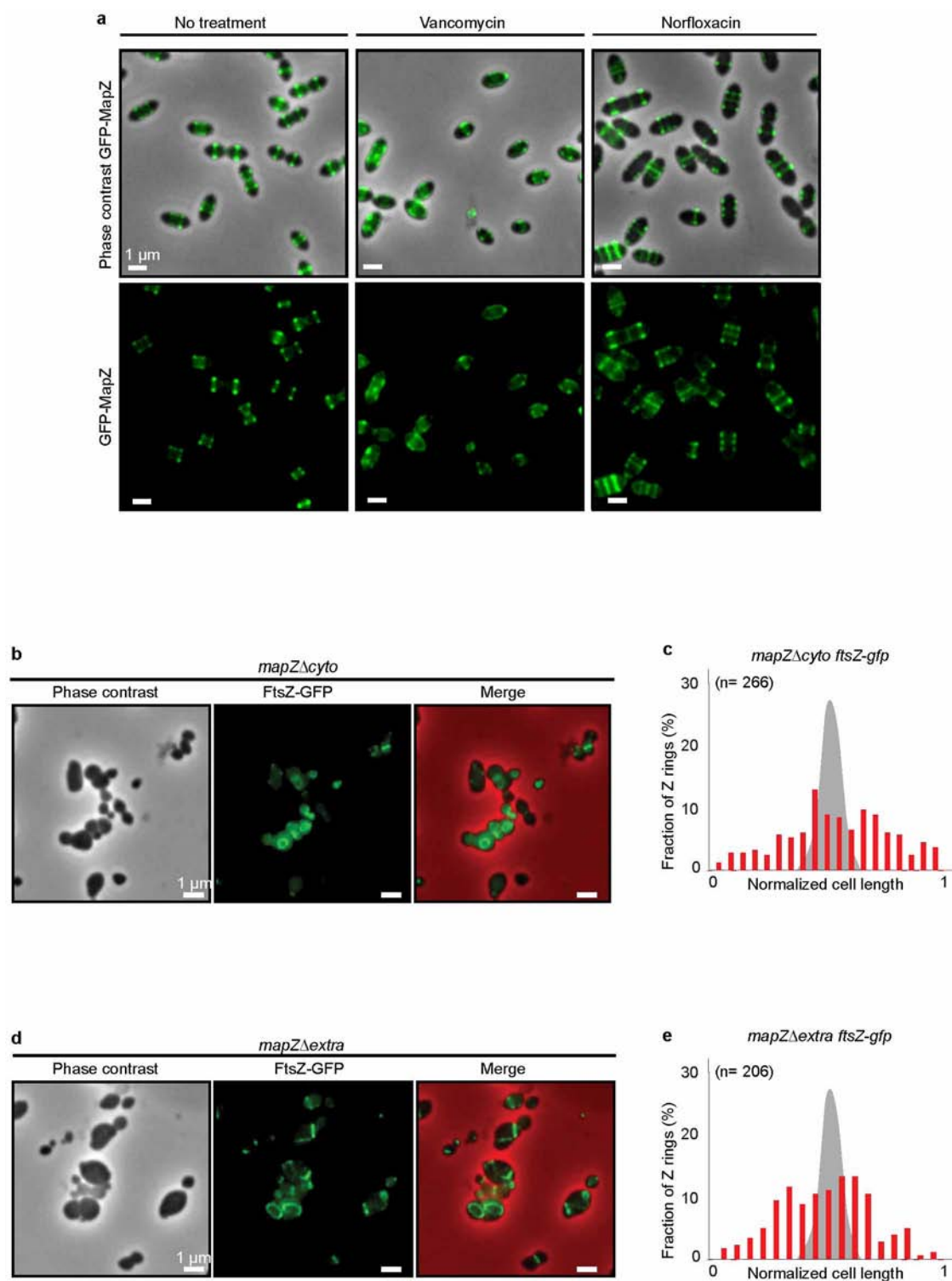
Extended Data Figure 5 | MapZ position depends on PG synthesis and FtsZ position depends on MapZ functionality. a, Localization of MapZ after inhibition of PG synthesis in wild-type pneumococcus. Microscopy images of GFP–MapZ in wild-type cells before (top), and 15 min after addition of vancomycin (middle) or norfloxacin (bottom). Vancomycin, which inhibits PG synthesis, impairs localization of GFP–MapZ, whereas norfloxacin, which inhibits topoisomerases, has no effect on MapZ septal localization. **b, c,** Localization of FtsZ–GFP in *mapZΔcyto* (**b**) and corresponding FtsZ–GFP ring positioning along the cell length normalized to 1 (**c**). **d, e,** Localization of FtsZ–GFP in *mapZΔextra* (**d**) and corresponding FtsZ–GFP ring positioning along the cell length normalized to 1 (**e**). Images are representative of experiments made in triplicate.
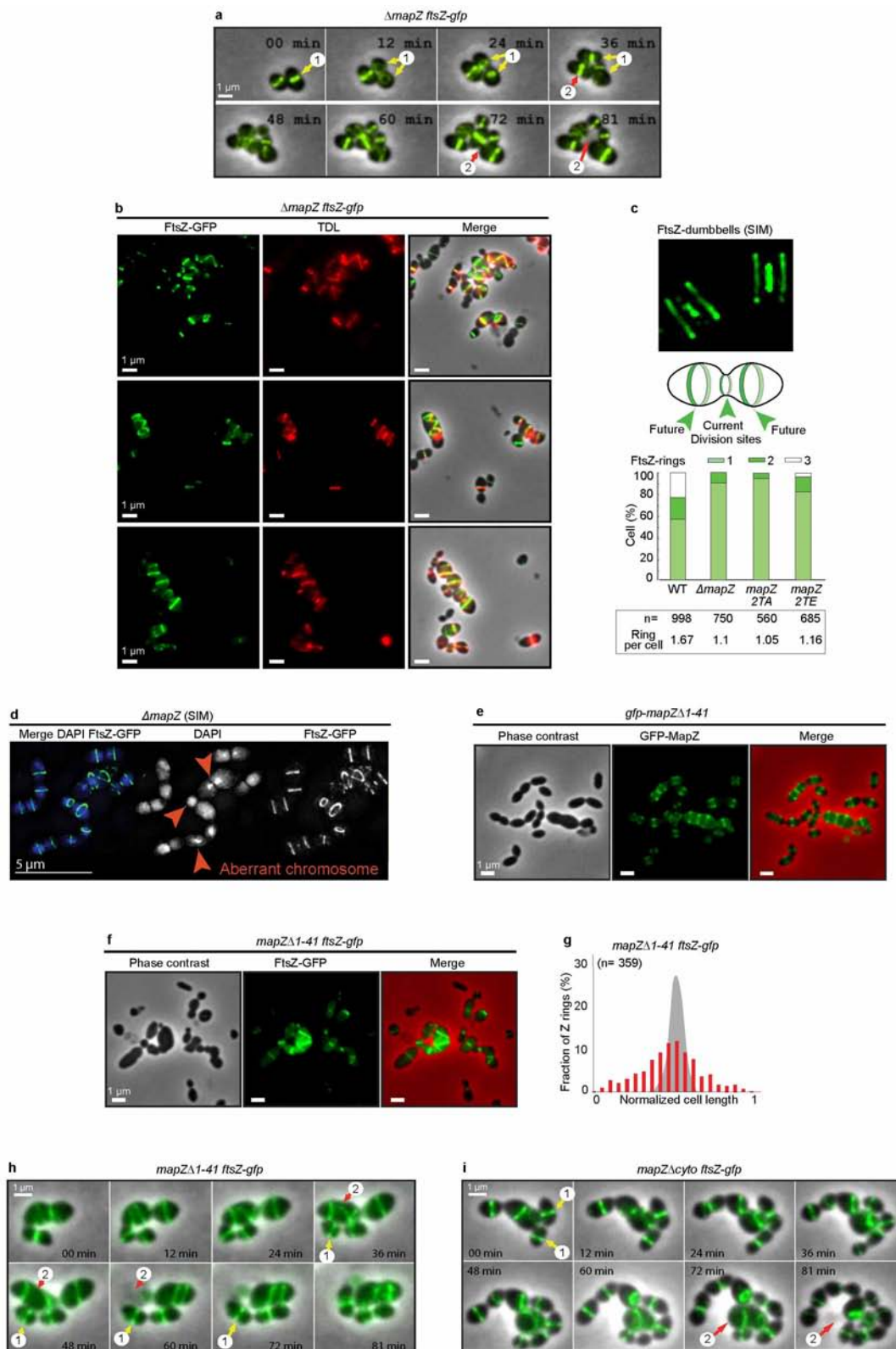
Extended Data Figure 6 | FtsZ is mispositioned in Δ*mapZ* cells and colocalizes with PG synthesis. a, Time-lapse images of FtsZ–GFP (green) dynamics in Δ*mapZ* cells. FtsZ polymers fail to position correctly even in cells with normal shape (arrow 1), resulting in asymmetric cell division or cell lysis (arrow 2) (stills correspond to Supplementary Video 5). **b**, Microscopy images showing colocalization of PG synthesis revealed by pulse labelling with TDL (red) and mispositioned FtsZ–GFP structures (green) in Δ*mapZ* cells. Three fields of view from three independent experiments are shown. **c**, 3D-SIM and schematic of FtsZ dumbbells with histograms of the cell ratios with 1, 2 or 3 rings. The average number of rings per cell is shown. **d**, 3D-SIM image of DAPI-stained DNA and FtsZ–GFP in Δ*mapZ* cells. **e**, Localization of GFP–MapZΔ(1–41) in *gfp-mapZΔ(1–41)*. **f**, Localization of FtsZ–GFP in *mapZΔ(1–41)*. **g**, Corresponding FtsZ–GFP ring positioning along the cell length normalized to 1. **h, i**, Time-lapse images of FtsZ–GFP (green) dynamics in *mapZΔ(1–41)* (**h**) and *mapZΔcyto* (**i**) cells. FtsZ mispositioning, even in cells with normal shape leads to asymmetric cell division (arrow 1) or cell lysis (arrow 2). Images are representative of experiments made in triplicate.
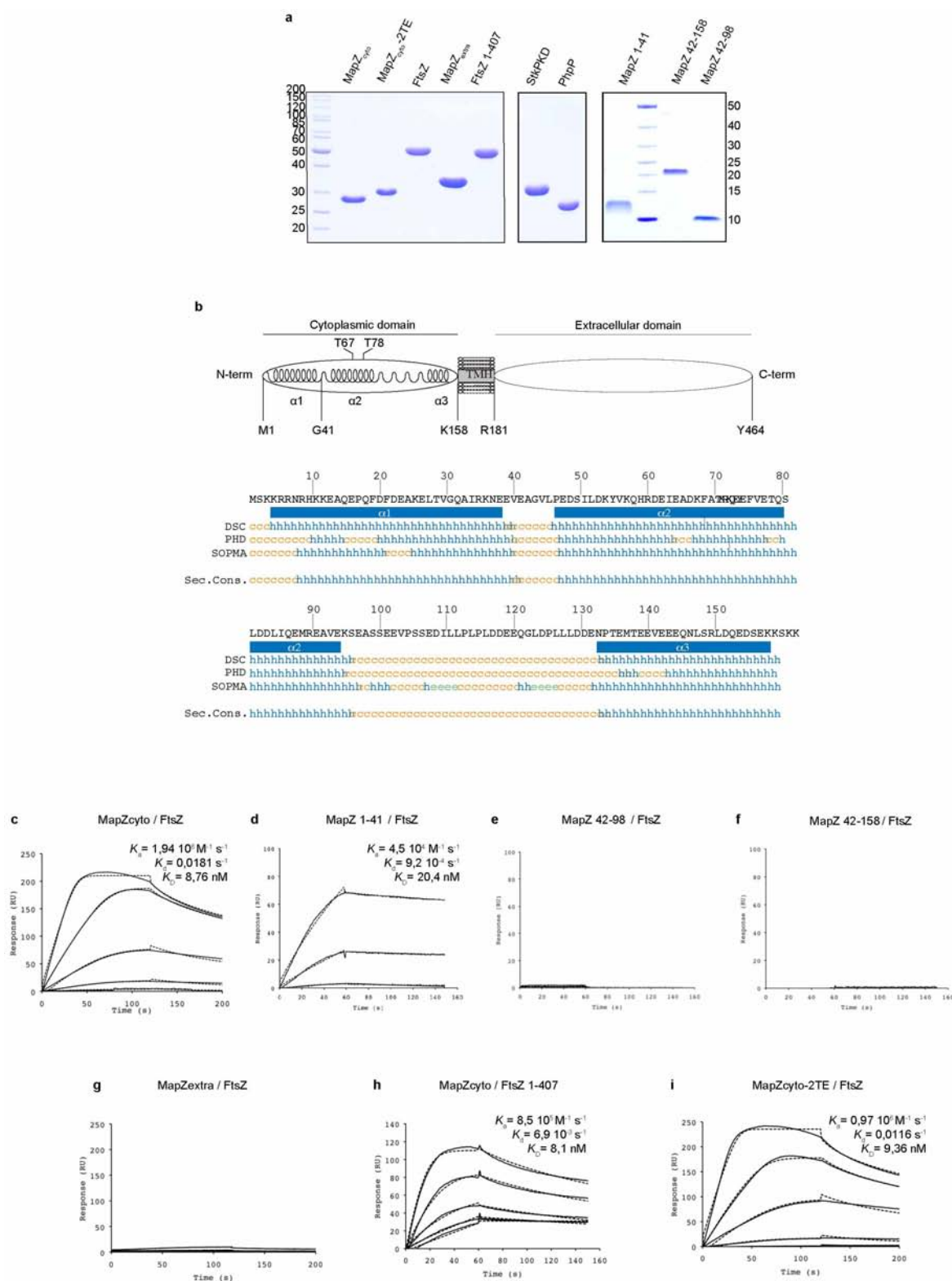
Extended Data Figure 7 | Purification of FtsZ and MapZ and analysis of the interactions between them. a, Purification of proteins used in surface plasmon resonance experiments. The MapZ cytoplasmic domain, MapZ-2TE cytoplasmic domain, FtsZ, MapZ extracellular domain, FtsZ$_{(1-407)}$ fragment (FtsZ deleted from the C-terminal α-helix), StkP cytoplasmic domain, PhpP, MapZ N-terminal peptide from Met 1 to Gly 41, MapZ peptide from Val 42 to Ser 98 and MapZ peptide from Val 42 to Lys 158 were overproduced in *E. coli* BL21 and analysed by SDS–PAGE. **b**, Schematic model of MapZ and secondary structure prediction of the cytoplasmic domain of MapZ. Secondary structure codes 'e', 'c' and 'h' indicate predicted α-helices (blue), random coils

(orange) and extended strands (green), respectively. **c–e**, Surface plasmon resonance analyses of interaction between FtsZ and MapZ. **c–i**, Full-length FtsZ (**c, d, e, f, g, i**) or FtsZ$_{1-407}$ (**h**) was covalently coupled to the surface of a CM5 sensorchip. Increasing amounts of either MapZ cytoplasmic domain (**c, h**), MapZ extracellular domain (**g**), MapZ-2TE (**i**) cytoplasmic domain, MapZ(1–41) (**d**), MapZ(42–98) (**e**) and MapZ(42–158) (**f**) peptides were injected onto the FtsZ- or FtsZ$_{(1-407)}$-coupled sensorship. RU, resonance units. The measurements were made in triplicate. The affinity (K_D), association (K_a) and dissociation constants (K_d) are indicated. Images are representative of experiments made in triplicate.
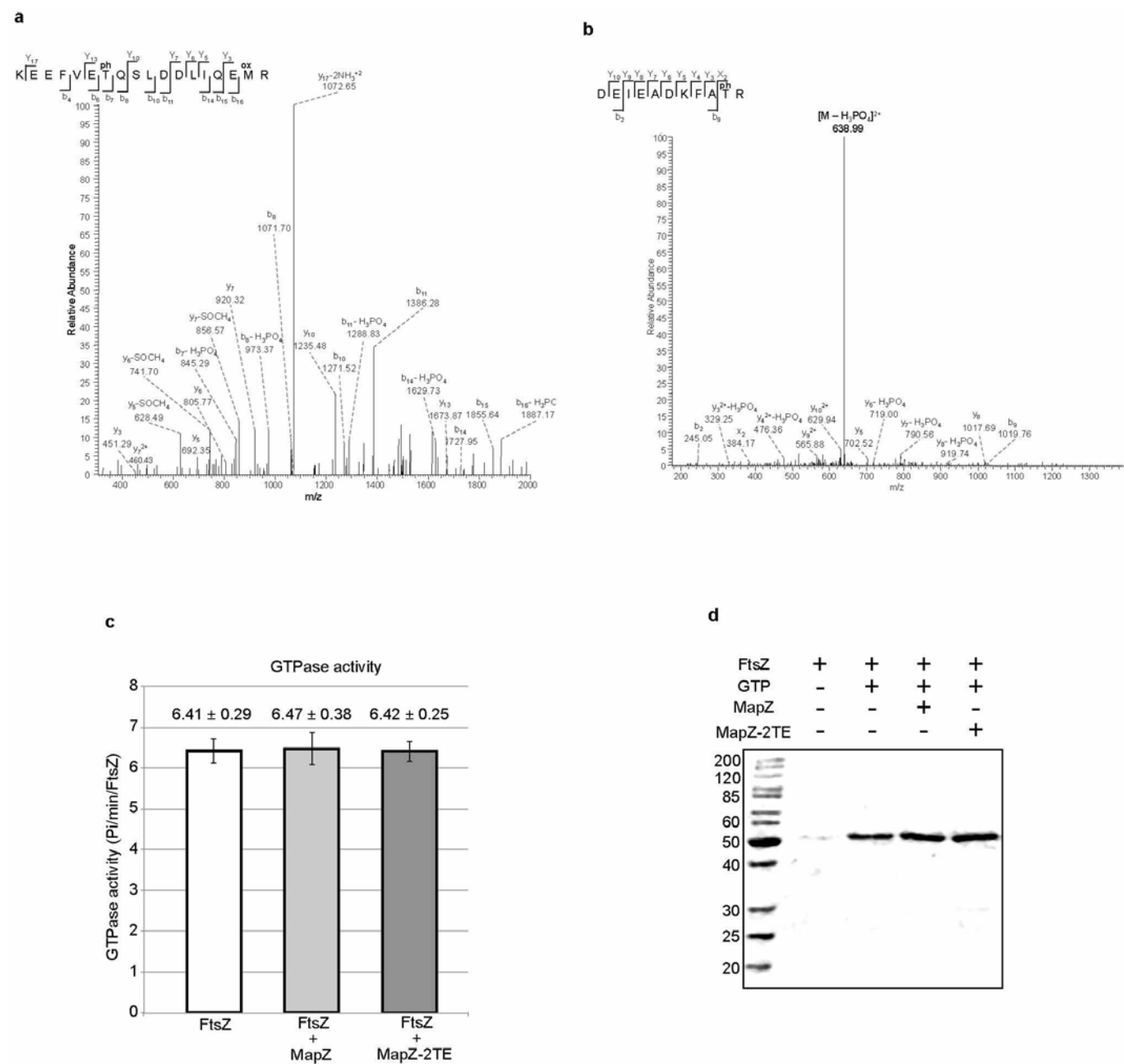
Extended Data Figure 8 | Analysis of MapZ *in vivo* phosphorylation and impact on FtsZ GTPase activity and polymerization. a, b, MapZ is phosphorylated on threonine 67 (**a**) and threonine 78 (**b**). The spectra show the fragmentation pattern of the phosphopeptides DEIEADKFAT(ph)R corresponding to amino acids 58–68 and KEEFVET(ph)QSLDDLIQEM(ox)R corresponding to amino acids 72–89. **c**, Influence of MapZ and MapZ-2TE cytoplasmic domains on FtsZ GTPase activity. Purified FtsZ was incubated with GTP either alone or in the presence of MapZ or MapZ-2TE cytoplasmic domains and free phosphate was revealed using malachite green colour development. Data are shown with s.d. for three independent experiments. **d**, FtsZ polymerization in the presence of MapZ$_{cyto}$, wild-type or mutated, cytoplasmic domains. FtsZ was incubated in the presence or absence of GTP and either MapZ or MapZ-2TE. The samples were then processed as described in Methods. Images are representative of experiments made in triplicate.

a *rfp-mapZ gfp-stkP*

b

c

	Bacterial Genus (#)*	Sequenced genomes (#)	Genomes with MapZ homologues (#)	(%)*
Bacteria	861	6,305	483	7.66
Other	740	4,629	0	0.00
Firmicutes	121	1,676	483	28.82
Other	83	330	0	0.00
Bacillales	23	567	1	0.18
Other	14	359	0	0.00
Bacillaceae	9	208	1	0.48
Other	8	20	0	0.00
Bacillus	1	188	1	0.53
Lactobacillales	15	779	482	61.87
Other	9	186	0	0.00
Enterococcaceae	4	169	70	41.42
Other	2	2	0	0.00
Enterococcus	1	165	68	41.21
Melissococcus	1	2	2	100.00
Streptococcaceae	2	424	412	97.17
Lactococcus	1	14	14	100.00
Streptococcus	1	410	398	97.07

* # for numbers and % for fractions

Extended Data Figure 9 | Interplay between MapZ and StkP and PhpP and conservation of MaZ in bacterial genomes. **a**, Simultaneous localization of GFP–StkP and RFP–MapZ in wild-type cells. Overlays between GFP (green), RFP (red) and phase contrast show that StkP locates at mid-cell while MapZ ring separation proceeds, as depicted in the summary diagram below. **b**, Dephosphorylation of MapZ by PhpP. MapZ cytoplasmic domain was phosphorylated by $StkP_{KD}$ and then incubated for various times (30 s to 10 min) with the protein phosphatase PhpP. MapZ dephosphorylation was analysed by autoradiography. **c**, Conservation analysis of *mapZ* homologues in 6,305 bacterial genomes. The left panel shows the taxonomy of the bacterial superkingdom. The right panel indicates the number of genera, the number of sequenced genomes, the number of genomes coding for MapZ homologous proteins and the percentage of genomes coding for MapZ homologous proteins. Images are representative of experiments made in triplicate.

Extended Data Table 1 | Strain viability and generation time

Strains	Viability[a] (%)	Generation time[b] (min)
WT	100	32 ± 3
ΔmapZ	70.3 ± 0,7	48 ± 2
mapZ-2TA	87 ± 2,7	31 ± 3
mapZ-2TE	95.8 ± 2,9	33 ± 2
mapZ+	97.9±1.1	35±2
ΔmapZ / Pzn-mapZ (no ZnCl$_2$)	78.6 ± 1.2	40± 2
ΔmapZ / Pzn-mapZ (0.1 mM ZnCl$_2$)	84 ± 2.1	39± 3
ΔmapZ / Pzn-mapZ (0.2 mM ZnCl$_2$)	89.5 ± 1.3	34.8 ±2
ftsZ-gfp	96.9 ± 4,3	34 ± 6
gfp-mapZ	95 ± 5	28 ± 3
mapZΔcyto	73.4 ± 2,7	46 ± 3
mapZΔ(1-41)	80.4± 1.3	36 ± 3
mapZΔextra	73.7 ± 0,8	49 ± 5

* Colony-forming units per millilitre (c.f.u. ml^{-1}) estimated by plating and normalized to that of wild-type strain. Data are shown with s.d. for three independent experiments.
† Time required for doubling of the optical density (OD$_{650\,nm}$) in liquid culture. Data are shown with s.d. for three independent experiments.

LETTER

Programmable RNA recognition and cleavage by CRISPR/Cas9

Mitchell R. O'Connell[1], Benjamin L. Oakes[1], Samuel H. Sternberg[2], Alexandra East-Seletsky[1], Matias Kaplan[3]†
& Jennifer A. Doudna[1,2,3,4]

The CRISPR-associated protein Cas9 is an RNA-guided DNA endonuclease that uses RNA–DNA complementarity to identify target sites for sequence-specific double-stranded DNA (dsDNA) cleavage[1–5]. In its native context, Cas9 acts on DNA substrates exclusively because both binding and catalysis require recognition of a short DNA sequence, known as the protospacer adjacent motif (PAM), next to and on the strand opposite the twenty-nucleotide target site in dsDNA[4–7]. Cas9 has proven to be a versatile tool for genome engineering and gene regulation in a large range of prokaryotic and eukaryotic cell types, and in whole organisms[8], but it has been thought to be incapable of targeting RNA[5]. Here we show that Cas9 binds with high affinity to single-stranded RNA (ssRNA) targets matching the Cas9-associated guide RNA sequence when the PAM is presented in *trans* as a separate DNA oligonucleotide. Furthermore, PAM-presenting oligonucleotides (PAMmers) stimulate site-specific endonucleolytic cleavage of ssRNA targets, similar to PAM-mediated stimulation of Cas9-catalysed DNA cleavage[7]. Using specially designed PAMmers, Cas9 can be specifically directed to bind or cut RNA targets while avoiding corresponding DNA sequences, and we demonstrate that this strategy enables the isolation of a specific endogenous messenger RNA from cells. These results reveal a fundamental connection between PAM binding and substrate selection by Cas9, and highlight the utility of Cas9 for programmable transcript recognition without the need for tags.

CRISPR–Cas immune systems must discriminate between self and nonself to avoid an autoimmune response[9]. In type I and II systems, foreign DNA targets that contain adjacent PAM sequences are targeted for degradation, whereas potential targets in CRISPR loci of the host do not contain PAMs and are avoided by RNA-guided interference complexes[3,5,6,10]. Single-molecule and bulk biochemical experiments showed that PAMs act both to recruit Cas9–guide-RNA (Cas9–gRNA) complexes to potential target sites and to trigger nuclease domain activation[7]. Cas9 from *Streptococcus pyogenes* recognizes a 5′-NGG-3′ PAM on the non-target (displaced) DNA strand[4,6], suggesting that PAM recognition may stimulate catalysis through allosteric regulation. Moreover, the HNH nuclease domain of Cas9, which mediates target-strand cleavage[4,5], is homologous to other HNH domains that cleave RNA substrates[11,12]. Based on the observations that single-stranded DNA (ssDNA) targets can be activated for cleavage by a separate PAMmer[7], and that similar HNH domains can cleave RNA, we wondered whether a similar strategy would enable Cas9 to cleave ssRNA targets in a programmable fashion (Fig. 1a).

Using *S. pyogenes* Cas9 and dual-guide RNAs (Methods), we performed *in vitro* cleavage experiments using a panel of RNA and DNA targets
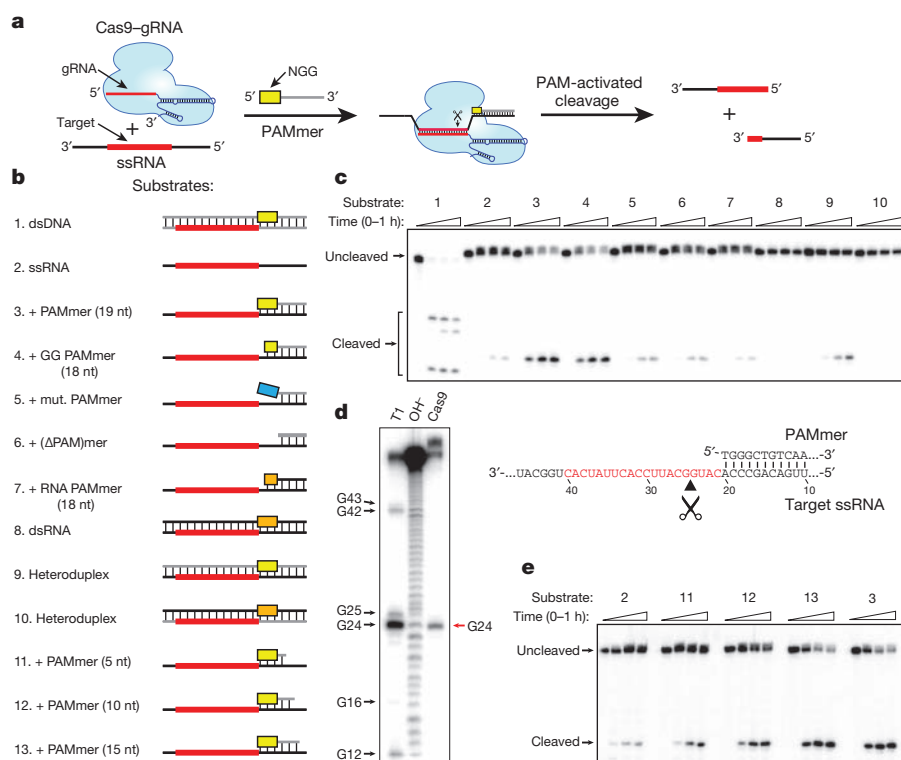
Figure 1 | RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM-presenting DNA oligonucleotide (PAMmer).
a, Schematic depicting the approach used to target ssRNA for programmable, sequence-specific cleavage. **b**, The panel of nucleic acid substrates examined in this study. Substrate elements are coloured as follows: DNA, grey; RNA, black; guide-RNA target sequence, red; DNA PAM, yellow; mutated DNA PAM, blue; RNA PAM, orange. The 18-nucleotide 'GG PAMmer' contains only a GG dinucleotide PAM sequence. nt, nucleotide. **c**, Representative cleavage assay for 5′-radiolabelled nucleic acid substrates using Cas9–gRNA, numbered as in **b**. **d**, Cas9–gRNA cleavage site mapping assay for substrate 3. T1 and OH⁻ denote RNase T1 and hydrolysis ladders, respectively; the sequence of the target ssRNA is shown at right. Sites of G cleavage by RNase T1 are shown at left. Site of Cas9 cleavage (G24) shown at right. **e**, Representative ssRNA cleavage assay in the presence of PAMmers of increasing length, numbered as in **b**.

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [2]Department of Chemistry, University of California, Berkeley, California 94720, USA. [3]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. †Present address: Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida 32611, USA.

(Fig. 1b and Extended Data Table 1). Deoxyribonucleotide-comprised PAMmers specifically activated Cas9 to cleave ssRNA (Fig. 1c), an effect that required a 5′-NGG-3′ or 5′-GG-3′ PAM. RNA cleavage was not observed using ribonucleotide-based PAMmers, suggesting that Cas9 may recognize the local helical geometry and/or deoxyribose moieties within the PAM. Consistent with this hypothesis, dsRNA targets were not cleavable and RNA–DNA heteroduplexes could only be cleaved when the non-target strand was composed of deoxyribonucleotides. Notably, we found that Cas9 cleaved the ssRNA target strand between positions 4 and 5 of the base-paired gRNA–target-RNA hybrid (Fig. 1d), in contrast to the cleavage between positions 3 and 4 observed for dsDNA[3–5]. This is probably due to subtle differences in substrate positioning. However, we did observe a significant reduction in the pseudo-first-order cleavage rate constant of PAMmer-activated ssRNA as compared to ssDNA[7] (Extended Data Fig. 1).

We hypothesized that PAMmer nuclease activation would depend on the stability of the hybridized PAMmer–ssRNA duplex and tested this by varying PAMmer length. As expected, ssRNA cleavage was lost when the predicted melting temperature for the duplex decreased below the temperature used in our experiments (Fig. 1e). In addition, large molar excesses of di- or tri-deoxyribonucleotides in solution were poor activators of Cas9 cleavage (Extended Data Fig. 2). Collectively, these data demonstrate that hybrid substrate structures composed of ssRNA and deoxyribonucleotide-based PAMmers that anneal upstream of the RNA target sequence can be cleaved efficiently by RNA-guided Cas9.

We investigated the binding affinity of catalytically inactive dCas9 (Cas9 (D10A;H840A))–gRNA for ssRNA targets with and without PAMmers using a gel mobility shift assay. Notably, whereas our previous results showed that ssDNA and PAMmer-activated ssDNA targets are bound with indistinguishable affinity[7], PAMmer-activated ssRNA targets were bound >500-fold tighter than ssRNA alone (Fig. 2a, b). A recent crystal structure of Cas9 bound to a ssDNA target revealed deoxyribose-specific van der Waals interactions between the protein and the DNA backbone[13],
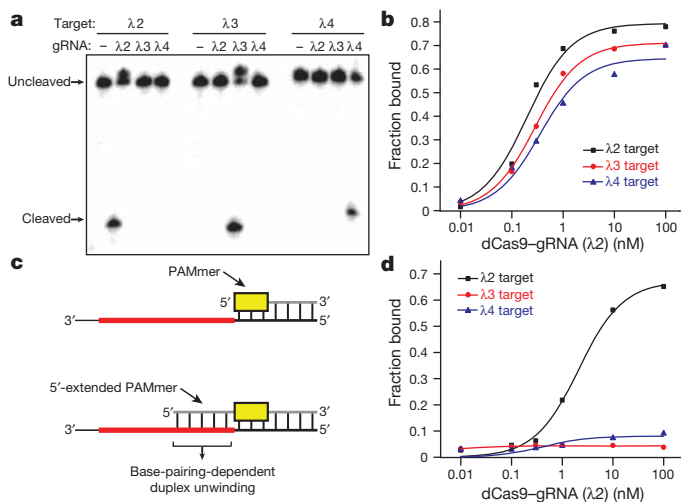
suggesting that energetic penalties associated with ssRNA binding must be attenuated by favourable compensatory binding interactions with the provided PAM. The equilibrium dissociation constant measured for a PAMmer–ssRNA substrate was within fivefold of that for dsDNA (Fig. 2b), and this high-affinity interaction again required a cognate deoxyribonucleotide-comprised 5′-GG-3′ PAM (Fig. 2a). Tight binding also scaled with PAMmer length (Fig. 2c), consistent with the cleavage data presented above.

It is known that Cas9 possesses an intrinsic affinity for RNA, but sequence specificity of the interaction had not been explored[5]. Thus, to verify the programmable nature of PAMmer-mediated ssRNA cleavage by Cas9–gRNA, we prepared three distinct guide RNAs (λ2, λ3 and λ4; each targeting 20-nucleotide sequences within λ2, λ3 and λ4 RNAs, respectively) and showed that their corresponding ssRNA targets could be efficiently cleaved using complementary PAMmers without any detectable cross-reactivity (Fig. 3a). This result indicates that complementary RNA–RNA base pairing is critical in these reactions. Notably however, dCas9 programmed with the λ2 guide RNA bound all three PAMmer–ssRNA substrates with similar affinity (Fig. 3b). This observation suggests that high-affinity binding in this case may not require correct base pairing between the guide RNA and the ssRNA target, particularly given the compensatory role of the PAMmer.

During dsDNA targeting by Cas9–gRNA, duplex melting proceeds directionally from the PAM and strictly requires the formation of complementary RNA–DNA base pairs to offset the energetic costs associated with dsDNA unwinding[7]. We therefore wondered whether binding specificity for ssRNA substrates would be recovered using PAMmers containing 5′-extensions that create a partially double-stranded target region requiring unwinding (Fig. 3c). We found that use of a 5′-extended PAMmer enabled dCas9 bearing the λ2 guide sequence to bind sequence-selectively to the λ2 PAMmer–ssRNA target. The λ3 and λ4 PAMmer–ssRNA targets were not recognized (Fig. 3d and Extended Data Fig. 3),

Figure 2 | dCas9–gRNA binds ssRNA targets with high affinity in the presence of PAMmers. a, Representative electrophoretic mobility shift assay for binding reactions with dCas9–gRNA and a panel of 5′-radiolabelled nucleic acid substrates, numbered as in Fig. 1b. **b**, Quantified binding data for substrates 1–4 from **a** fitted with standard binding isotherms. Measured dissociation constants from three independent experiments (mean ± s.d.) were 0.036 ± 0.003 nM (substrate 1), >100 nM (substrate 2), 0.20 ± 0.09 nM (substrate 3) and 0.18 ± 0.07 nM (substrate 4). **c**, Relative binding data for 1 nM dCas9–gRNA and 5′-radiolabelled ssRNA with a panel of different PAMmers. The data are normalized to the amount of binding observed at 1 nM dCas9–gRNA with a 19-nucleotide (nt) PAMmer; error bars represent the standard deviation from three independent experiments.

Figure 3 | 5′-extended PAMmers are required for specific target ssRNA binding. a, Cas9 programmed with either λ2-, λ3- or λ4-targeting gRNAs exhibits sequence-specific cleavage of 5′-radiolabelled λ2, λ3 and λ4 target ssRNAs, respectively, in the presence of cognate PAMmers. **b**, dCas9 programmed with a λ2-targeting gRNA exhibits similar binding affinity to λ2, λ3 and λ4 target ssRNAs in the presence of cognate PAMmers. Dissociation constants from three independent experiments (mean ± s.d.) were 0.20 ± 0.09 nM (λ2), 0.33 ± 0.14 nM (λ3) and 0.53 ± 0.21 nM (λ4). **c**, Schematic depicting the approach used to restore gRNA-mediated ssRNA binding specificity, which involves 5′-extensions to the PAMmer that cover part or all of the target sequence. **d**, dCas9 programmed with a λ2-targeting gRNA specifically binds the λ2 ssRNA but not λ3 and λ4 ssRNAs in the presence of complete 5′-extended PAMmers. Dissociation constants from three independent experiments (mean ± s.d.) were 3.3 ± 1.2 nM (λ2) and >100 nM (λ3 and λ4).

although we did observe a tenfold reduction in overall ssRNA substrate binding affinity. By systematically varying the length of the 5′ extension, we found that PAMmers containing 2–8 additional nucleotides upstream of the 5′-NGG-3′ offer an optimal compromise between gains in binding specificity and concomitant losses in binding affinity and cleavage efficiency (Extended Data Fig. 4).

Next we investigated whether nuclease activation by PAMmers requires base pairing between the 5′-NGG-3′ and corresponding nucleotides on the ssRNA. Prior studies have shown that DNA substrates containing a cognate PAM that is mismatched with the corresponding nucleotides on the target strand are cleaved as efficiently as a fully base-paired PAM[4]. This could enable targeting of RNA while precluding binding or cleavage of corresponding genomic DNA sites lacking PAMs (Fig. 4a). To test this possibility, we first demonstrated that Cas9–gRNA cleaves PAMmer–ssRNA substrates regardless of whether or not the PAM is base paired (Fig. 4b, c). When Cas9–RNA was incubated with both a PAMmer–ssRNA substrate and the corresponding dsDNA template containing a cognate PAM, both targets were cleaved. In contrast, when a dsDNA target lacking a PAM was incubated together with a PAMmer–ssRNA substrate bearing a mismatched 5′-NGG-3′ PAM, Cas9–gRNA selectively targeted the ssRNA for cleavage (Fig. 4c). The same result was obtained using a mismatched PAMmer with a 5′ extension (Fig. 4c), demonstrating that this general strategy enables the specific targeting

of RNA transcripts while effectively eliminating any targeting of their corresponding dsDNA template loci.

We next explored whether Cas9-mediated RNA targeting could be applied in tagless transcript isolation from HeLa cells (Fig. 4d). The immobilization of Cas9 on a solid-phase resin is described in Methods (see also Extended Data Fig. 5). As a proof of concept, we first isolated *GAPDH* mRNA from HeLa total RNA using biotinylated dCas9, gRNAs and PAMmers (Extended Data Table 2) that target four non-PAM-adjacent sequences within exons 5–7 (Fig. 4e). We observed a substantial enrichment of *GAPDH* mRNA relative to control β-actin mRNA by northern blot analysis, but saw no enrichment using a non-targeting gRNA or dCas9 alone (Fig. 4f).

We then used this approach to isolate endogenous *GAPDH* transcripts from HeLa cell lysate under physiological conditions. In initial experiments, we found that Cas9–gRNA captured two *GAPDH*-specific RNA fragments rather than the full-length mRNA (Fig. 4g). Based on the sizes of these bands, we hypothesized that RNA–DNA heteroduplexes formed between the mRNA and PAMmer were cleaved by cellular RNase H. Previous studies have shown that modified DNA oligonucleotides can abrogate RNase H activity[14], and therefore we investigated whether Cas9 would tolerate chemical modifications to the PAMmer. We found that a wide range of modifications (locked nucleic acids, 2′-OMe and 2′-F ribose moieties) still enabled PAMmer-mediated nuclease
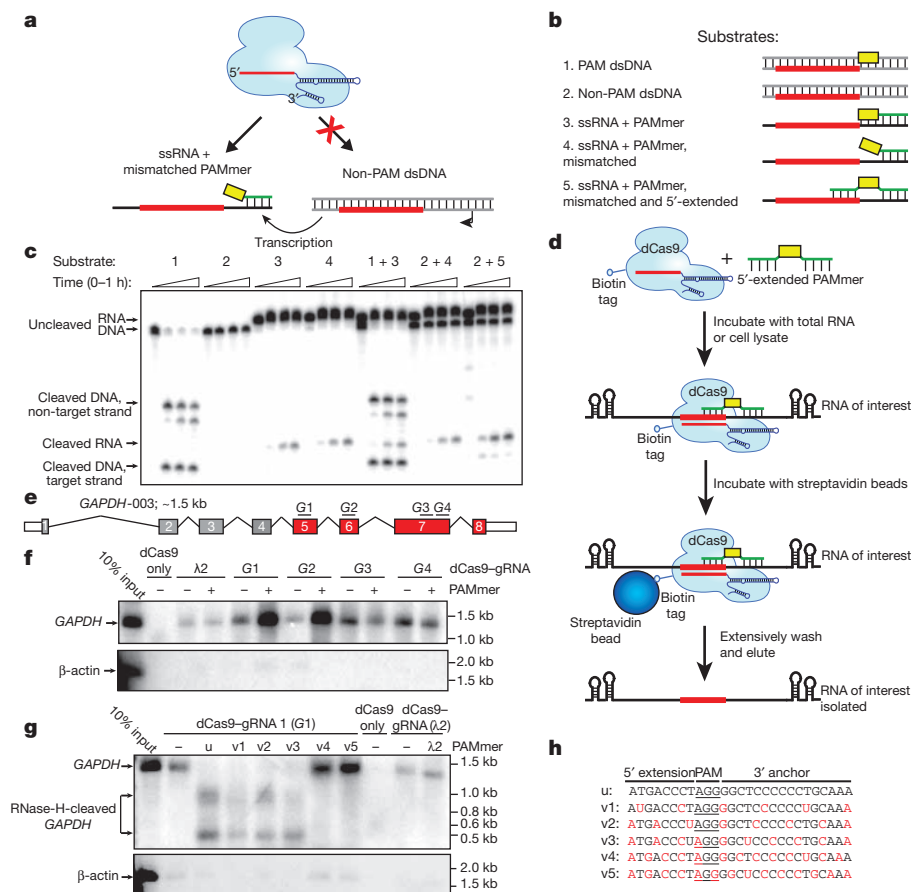
Figure 4 | RNA-guided Cas9 can target non-PAM sites on ssRNA and isolate *GAPDH* mRNA from HeLa cells in a tagless manner. a, Schematic of the approach designed to avoid cleavage of template DNA by targeting non-PAM sites in the ssRNA target. **b**, The panel of nucleic acid substrates tested in **c**. **c**, Cas9–gRNA cleaves ssRNA targets with equal efficiency when the 5′-NGG-3′ of the PAMmer is mismatched with the ssRNA. This strategy enables selective cleavage of ssRNA in the presence of non-PAM target dsDNA. **d**, Schematic of the dCas9 RNA pull-down experiment. **e**, *GAPDH* mRNA transcript isoform 3 (*GAPDH*-003) shown schematically, with exons common to all *GAPDH* protein-coding transcripts in red and gRNA/PAMmer targets

G1–G4 indicated. kb, kilobase pairs. **f**, Northern blot showing that gRNAs and corresponding 5′-extended PAMmers enable tagless isolation of *GAPDH* mRNA from HeLa total RNA; β-actin mRNA is shown as a control. **g**, Northern blot showing tagless isolation of *GAPDH* mRNA from HeLa cell lysate with varying 2′-OMe-modified PAMmers. RNase H cleavage is abrogated with v4 and v5 PAMmers; β-actin mRNA is shown as a control. u, unmodified PAMmer (*G1*). v1-v5, increasingly 2′-OMe-modified PAMmers (*G1*), see **g** for PAMmer sequences. **h**, Sequences of unmodified and modified *GAPDH* PAMmers used in **g**; 2′-OMe-modified nucleotides are shown in red.

activation (Extended Data Fig. 6). Furthermore, by varying the pattern of 2′-OMe modifications in the PAMmer, we could completely eliminate RNase-H-mediated cleavage during the pull-down and successfully isolate intact *GAPDH* mRNA (Fig. 4g, h). Notably, we consistently observed specific isolation of *GAPDH* mRNA in the absence of any PAMmer, albeit with lower efficiency, suggesting that Cas9–gRNA can bind to *GAPDH* mRNA through direct RNA–RNA hybridization (Fig. 4f, g and Extended Data Fig. 7). These experiments demonstrate that RNA-guided Cas9 can be used to purify endogenous untagged RNA transcripts. In contrast to current oligonucleotide-mediated RNA-capture methods, this approach works well under physiological salt conditions and does not require crosslinking or large sets of biotinylated probes[15–17].

Here we have demonstrated the ability to re-direct the dsDNA targeting capability of CRISPR/Cas9 for RNA-guided ssRNA binding and/or cleavage (which we now denote RCas9, an RNA-targeting Cas9). Programmable RNA recognition and cleavage has the potential to transform the study of RNA function, much as site-specific DNA targeting is changing the landscape of genetic and genomic research[8] (Extended Data Fig. 8). Although certain engineered proteins such as PPR proteins and Pumilio/FBF (PUF) repeats show promise as platforms for sequence-specific RNA targeting[18–22], these strategies require re-designing the protein for every new RNA sequence of interest. While RNA interference has proven useful for manipulating gene regulation in certain organisms[23], there has been a strong motivation to develop orthogonal nucleic-acid-based RNA recognition systems, such as the CRISPR/Cas Type III-B Cmr complex[24–28] and the atypical Cas9 from *Francisella novicida*[29,30]. In contrast to these systems, the molecular basis for RNA recognition by RCas9 is now clear and requires only the design and synthesis of a matching gRNA and complementary PAMmer. The ability to recognize endogenous RNAs within complex mixtures with high affinity and in a programmable manner paves the way for direct transcript detection, analysis and manipulation without the need for genetically encoded affinity tags.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
3. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
4. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
6. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol* **155**, 733–740 (2009).
7. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
8. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nature Methods* **10**, 957–963 (2013).
9. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
10. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46**, 606–615 (2012).
11. Pommer, A. J. *et al.* Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* **314**, 735–749 (2001).
12. Hsia, K. C. *et al.* DNA binding and degradation by the HNH protein ColE7. *Structure* **12**, 205–214 (2004).
13. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
14. Wu, H. J., Lima, W. F. & Crooke, S. T. Properties of cloned and expressed human RNase H1. *J. Biol. Chem.* **274**, 28270–28278 (1999).
15. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341** (2013).
16. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
17. Simon, M. D. *et al.* The genomic binding sites of a noncoding RNA. *Proc. Natl Acad. Sci. USA* **108**, 20497–20502 (2011).
18. Mackay, J. P., Font, J. & Segal, D. J. The prospects for designer single-stranded RNA-binding proteins. *Nature Struct. Mol. Biol.* **18**, 256–261 (2011).
19. Filipovska, A. & Rackham, O. Designer RNA-binding proteins: new tools for manipulating the transcriptome. *RNA Biol.* **8**, 978–983 (2011).
20. Wang, Y., Wang, Z. & Tanaka Hall, T. M. Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J.* **280**, 3755–3767 (2013).
21. Yin, P. *et al.* Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168–171 (2013).
22. Yagi, Y., Nakamura, T. & Small, I. The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* **78**, 772–782 (2014).
23. Kim, D. H. & Rossi, J. J. RNAi mechanisms and applications. *Biotechniques* **44**, 613–616 (2008).
24. Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell* **139**, 945–956 (2009).
25. Hale, C. R. *et al.* Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* **45**, 292–302 (2012).
26. Staals, R. H. J. *et al.* Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol. Cell* **52**, 135–145 (2013).
27. Spilman, M. *et al.* Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol. Cell* **52**, 146–152 (2013).
28. Terns, R. M. & Terns, M. P. CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet.* **30**, 111–118 (2014).
29. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254–257 (2013).
30. Sampson, T. R. & Weiss, D. S. Exploiting CRISPR/Cas systems for biotechnology. *Bioessays* **36**, 34–38 (2014).

Author Contributions M.R.O. and S.H.S. conceived the project. M.R.O., B.L.O., S.H.S., A.E.-S. and M.K. conducted experiments. All authors discussed the data, and M.R.O., S.H.S., B.L.O. and J.A.D. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.D. (doudna@berkeley.edu).

METHODS

Cas9 and nucleic acid preparation. Wild-type Cas9 and catalytically inactive dCas9 (Cas9(D10A;H840A)) from *S. pyogenes* were purified as previously described[4]. Forty-two-nucleotide crRNAs were either ordered synthetically (Integrated DNA Technologies) or transcribed *in vitro* with T7 polymerase using single-stranded DNA templates, as described[31]. Using the previously described numbering scheme[4], tracrRNA was transcribed *in vitro* and contained nucleotides 15–87. Single-guide RNAs (sgRNAs) targeting λ-RNAs were transcribed *in vitro* from linearized plasmids and contain full-length crRNA and tracrRNA connected via a GAAA tetra-loop insertion. *GAPDH* mRNA-targeting sgRNAs were transcribed *in vitro* from dsDNA PCR products based on an optimized sgRNA design[32]. Target ssRNAs (55–56 nucleotides) were transcribed *in vitro* using single-stranded DNA templates. Sequences of all nucleic acid substrates used in this study can be found in Extended Data Tables 1 and 2.

All RNAs were purified using 10–15% denaturing polyacrylamide gel electrophoresis (PAGE). Duplexes of crRNA and tracrRNA were prepared by mixing equimolar concentrations of each RNA in hybridization buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM MgCl$_2$), heating to 95 °C for 30 s and slow cooling. Fully double-stranded DNA/RNA substrates (substrates 1, 8–10 in Fig. 1 and substrates 1 and 2 in Fig. 4) were prepared by mixing equimolar concentrations of each nucleic acid strand in hybridization buffer, heating to 95 °C for 30 s, and slow cooling. RNA, DNA and chemically modified PAMmers were synthesized commercially (Intergrated DNA Technologies). DNA and RNA substrates were 5′-radiolabelled using [γ-^{32}P]ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs). Double-stranded DNA and dsRNA substrates (Figs 1c and 4c) were 5′-radiolabelled on both strands, whereas only the target ssRNA was 5′-radiolabelled in other experiments.

Cleavage assays. Cas9–gRNA complexes were reconstituted before cleavage experiments by incubating Cas9 and the crRNA–tracrRNA duplex for 10 min at 37 °C in reaction buffer (20 mM Tris-HCl, pH 7.5, 75 mM KCl, 5 mM MgCl$_2$, 1 mM dithiothreitol (DTT), 5% glycerol). Cleavage reactions were conducted at 37 °C and contained ~1 nM 5′-radiolabelled target substrate, 100 nM Cas9–RNA, and 100 nM PAMmer, where indicated. Aliquots were removed at each time point and quenched by the addition of RNA gel-loading buffer (95% deionized formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 50 mM EDTA (pH 8.0), 0.025% (w/v) SDS). Samples were boiled for 10 min at 95 °C before being resolved by 12% denaturing PAGE. Reaction products were visualized by phosphorimaging and quantified with ImageQuant (GE Healthcare).

RNA cleavage site mapping. A hydrolysis ladder (OH⁻) was obtained by incubating ~25 nM 5′-radiolabelled λ2 target ssRNA in hydrolysis buffer (25 mM CAPS (*N*-cyclohexyl-3-aminopropanesulphonic acid), pH 10.0, 0.25 mM EDTA) at 95 °C for 10 min, before quenching on ice. An RNase T1 ladder was obtained by incubating ~25 nM 5′-radiolabelled λ2 target ssRNA with 1 U RNase T1 (New England Biolabs) for 5 min at 37 °C in RNase T1 buffer (20 mM sodium citrate, pH 5.0, 1 mM EDTA, 2 M urea, 0.1 mg ml^{-1} yeast transfer RNA). The reaction was quenched by phenol/chloroform extraction before adding RNA gel-loading buffer. All products were resolved by 15% denaturing PAGE.

Electrophoretic mobility shift assays. In order to avoid dissociation of the Cas9–gRNA complex at low concentrations during target ssRNA binding experiments, binding reactions contained a constant excess of dCas9 (300 nM), increasing concentrations of sgRNA, and 0.1–1 nM of target ssRNA. The reaction buffer was supplemented with 10 μg ml^{-1} heparin in order to avoid non-specific association of *apo*-dCas9 with target substrates[7]. Reactions were incubated at 37 °C for 45 min before being resolved by 8% native PAGE at 4 °C (0.5 × TBE buffer with 5 mM MgCl$_2$). RNA and DNA were visualized by phosphorimaging, quantified with ImageQuant (GE Healthcare), and analysed with Kaleidagraph (Synergy Software).

Cas9 biotin labelling. To ensure specific labelling at a single residue on Cas9, two naturally occurring cysteine residues were mutated to serine (C80S and C574S) and a cysteine point mutant was introduced at residue Met 1. To attach the biotin moiety, 10 μM wild-type Cas9 or dCas9 was reacted with a 50-fold molar excess of EZ-Link Maleimide-PEG2-Biotin (Thermo Scientific) at 25 °C for 2 h. The reaction was quenched by the addition of 10 mM DTT, and unreacted Maleimide-PEG2-Biotin was removed using a Bio-Gel P-6 column (Bio-Rad). Labelling was verified using a streptavidin bead binding assay, where 8.5 pmol of biotinylated Cas9 or non-biotinylated Cas9 was mixed with either 25 μl streptavidin-agarose (Pierce Avidin Agarose; Thermo Scientific) or 25 μl streptavidin magnetic beads (Dynabeads MyOne Streptavidin C1; Life Technologies). Samples were incubated in Cas9 reaction buffer at room temperature for 30 min, followed by three washes with Cas9 reaction buffer and elution in boiling SDS–PAGE loading buffer. Elutions were analysed using SDS–PAGE. Cas9 M1C biotinylation was also confirmed using mass spectroscopy performed in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley. Samples of intact Cas9 proteins were analysed using an Agilent 1200 liquid chromatograph equipped with a Viva C8 (100 mm × 1.0 mm, 5 μm particles, Restek) analytical column and connected in-line with an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Mass spectra were recorded in the positive ion mode. Mass spectral deconvolution was performed using ProMass software (Novatia).

***GAPDH* mRNA pull-down.** HeLa-S3 cell lysates were prepared as previously described[33]. Total RNA was isolated from HeLa-S3 cells using Trizol reagent according to the manufacturer's instructions (Life Technologies). Cas9–sgRNA complexes were reconstituted before pull-down experiments by incubating a twofold molar excess of Cas9 with sgRNA for 10 min at 37 °C in reaction buffer. HeLa total RNA (40 μg) or HeLa lysate (~5 × 10^6 cells) was added to reaction buffer with 40 U RNasin (Promega), PAMmer (5 μM) and the biotin-dCas9 (50 nM)–sgRNA (25 nM) in a total volume of 100 μl and incubated at 37 °C for 1 h. This mixture was then added to 25 μl magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; Life Technologies) pre-equilibrated in reaction buffer and agitated at 4 °C for 2 h. Beads were then washed six times with 300 μl wash buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM MgCl$_2$, 0.1% Triton X-100, 5% glycerol, 1 mM DTT, 10 μg ml^{-1} heparin). Immobilized RNA was eluted by heating beads at 70 °C in the presence of DEPC-treated water and a phenol/chloroform mixture. Eluates were then treated with an equal volume of glyoxal loading dye (Life Technologies) and heated at 50 °C for 1 h before separation via 1% BPTE agarose gel (30 mM Bis-Tris, 10 mM PIPES, 10 mM EDTA, pH 6.5). Northern blot transfers were carried out as previously described[34]. Following transfer, membranes were crosslinked using UV radiation and incubated in pre-hybridization buffer (UltraHYB Ultrasensitive Hybridization Buffer; Life Technologies) for 1 h at 46 °C before hybridization. Radioactive northern probes were synthesized using random priming of *GAPDH* and β-actin partial cDNAs (for cDNA primers, see Extended Data Table 2) in the presence of [α-^{32}P]dATP (PerkinElmer), using a Prime-It II Random Primer Labelling kit (Agilent Technologies). Hybridization was carried out for 3 h in pre-hybridization buffer at 46 °C followed by two washes with 2 × SSC (300 mM NaCl, 30 mM trisodium citrate, pH 7, 0.5% (w/v) SDS) for 15 min at 46 °C. Membranes were imaged using a phosphorscreen.

31. Sternberg, S. H., Haurwitz, R. E. & Doudna, J. A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**, 661–672 (2012).
32. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
33. Lee, H. Y. *et al.* RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl Acad. Sci. USA* **110**, 5416–5421 (2013).
34. Chomczynski, P. One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Anal. Biochem.* **201**, 134–139 (1992).
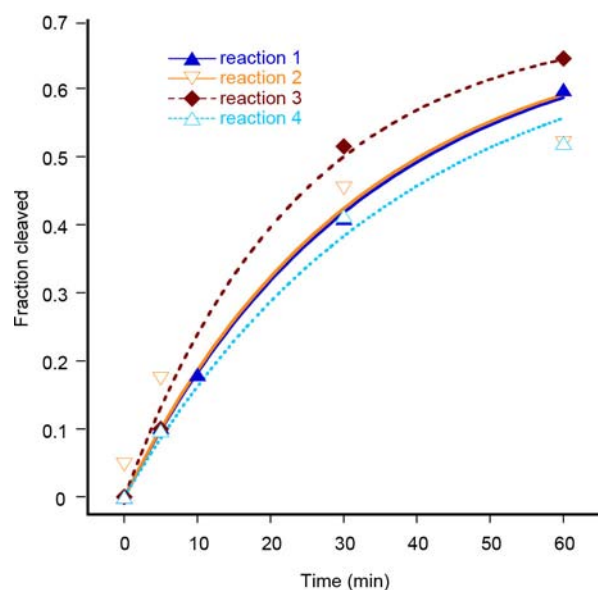
Extended Data Figure 1 | Quantified data for cleavage of ssRNA by Cas9–gRNA in the presence of a 19-nucleotide PAMmer. Cleavage assays were conducted as described in the Methods, and the quantified data were fitted with single-exponential decays. Results from four independent experiments yielded an average apparent pseudo-first-order cleavage rate constant (mean ± s.d.) of $0.032 \pm 0.007\,\text{min}^{-1}$. This is slower than the rate constant determined previously for ssDNA in the presence of the same 19-nucleotide PAMmer ($7.3 \pm 3.2\,\text{min}^{-1}$)[7].
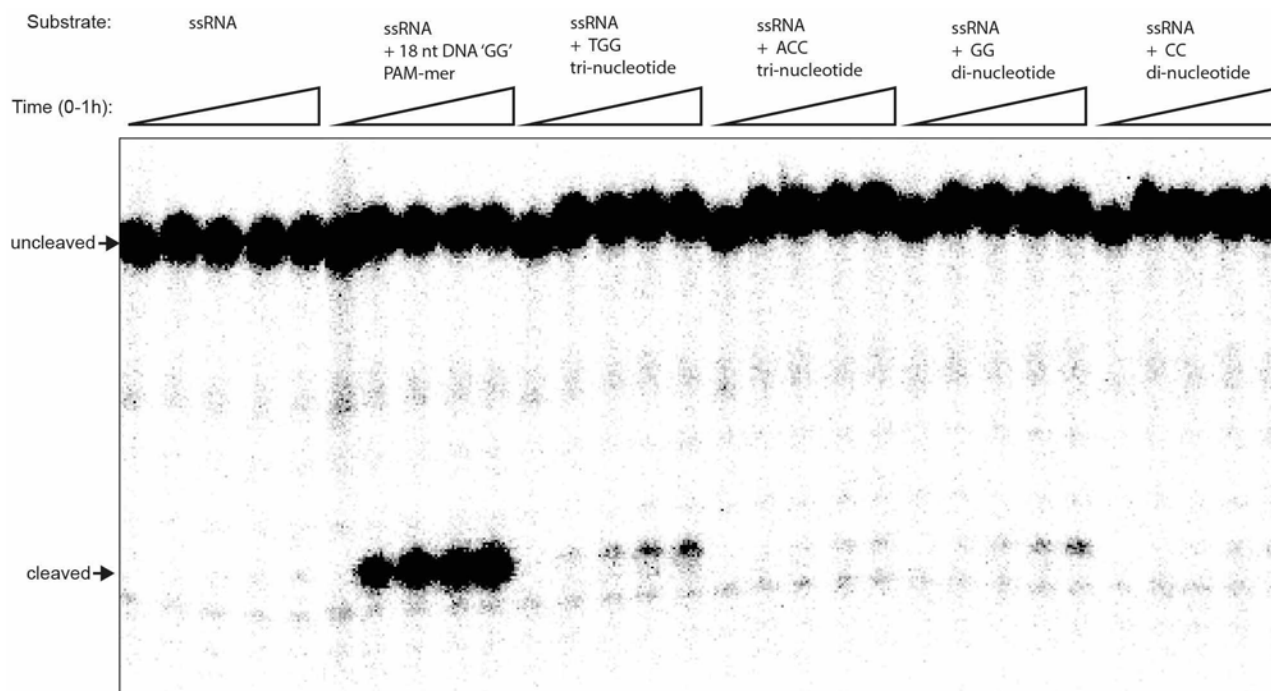
Extended Data Figure 2 | RNA cleavage is marginally stimulated by di- and tri-deoxyribonucleotide PAMmers. Cleavage reactions contained ~1 nM 5′-radiolabelled target ssRNA and no PAMmer (left), 100 nM 18-nt PAMmer (second from left), or 1 mM of the indicated di- or tri-nucleotide (remaining lanes). Reaction products were resolved by 12% denaturing polyacrylamide gel electrophoresis (PAGE) and visualized by phosphorimaging.
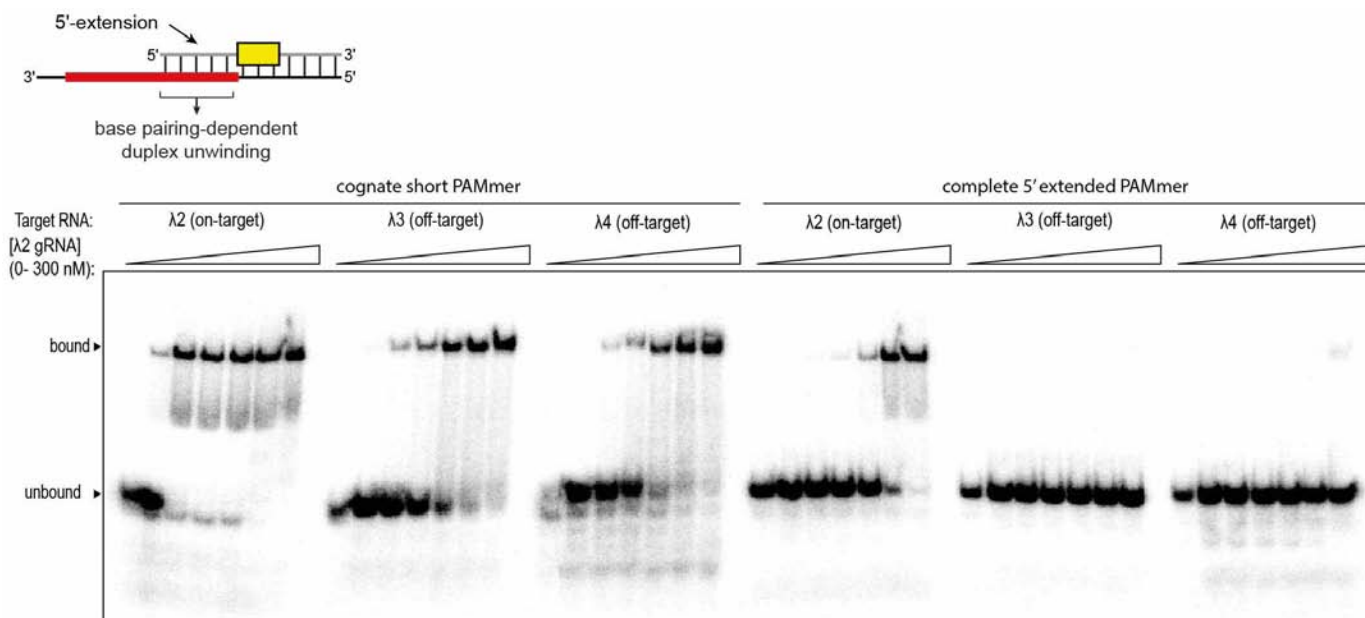
Extended Data Figure 3 | Representative binding experiment demonstrating guide-specific ssRNA binding with 5′-extended PAMmers. Gel shift assays were conducted as described in the Methods. Binding reactions contained Cas9 programmed with λ2 gRNA and either λ2 (on-target), λ3 (off-target) or λ4 (off-target) ssRNA in the presence of short cognate PAMmers or cognate PAMmers with complete 5′-extensions, as indicated. The presence of a cognate 5′-extended PAM-mer abrogates off-target binding. Three independent experiments were conducted to produce the data shown in Fig. 3b, d.
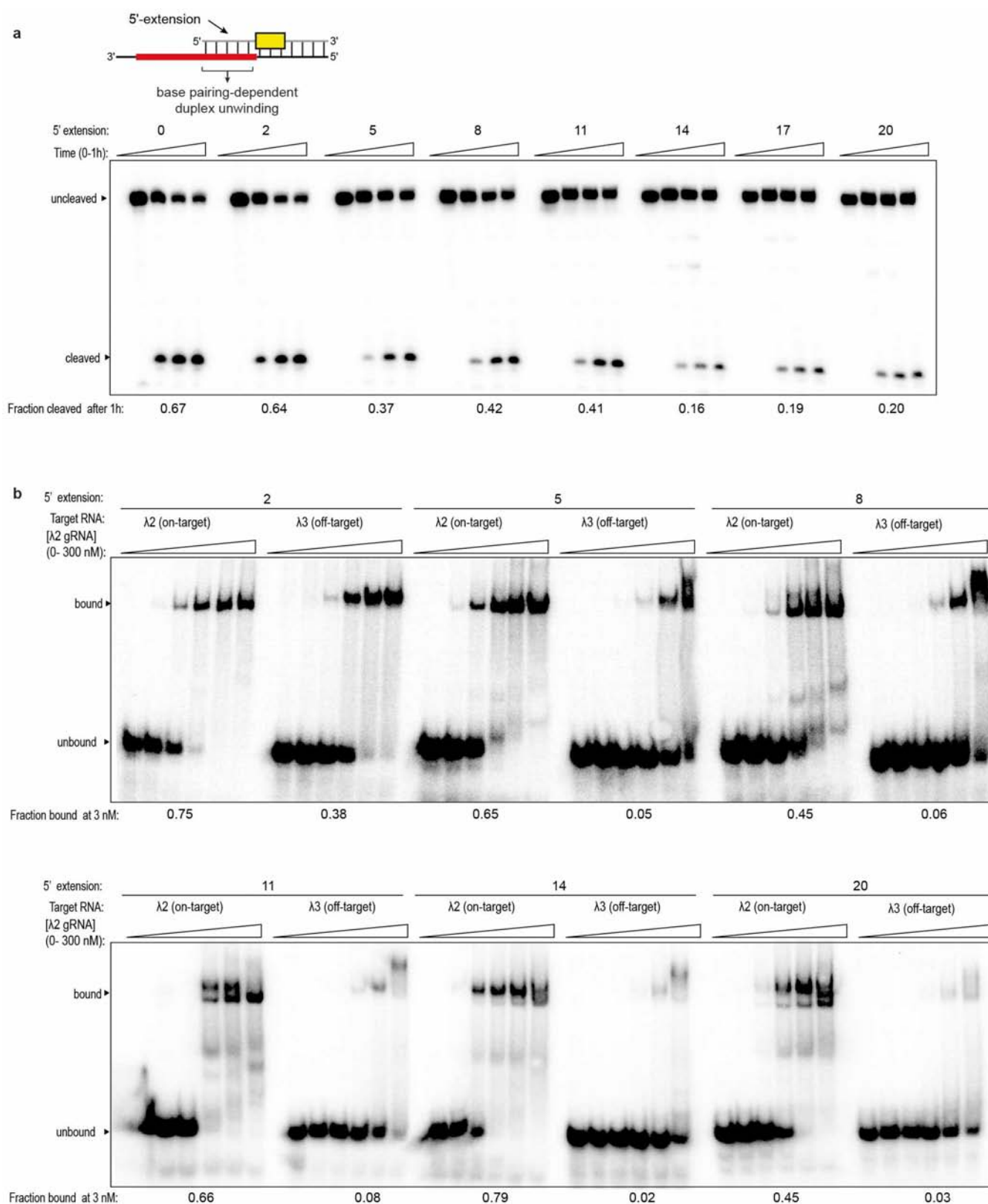
Extended Data Figure 4 | Exploration of RNA cleavage efficiencies and binding specificity using PAMmers with variable 5′-extensions. a, Cleavage assays were conducted as described in Methods. Reactions contained Cas9 programmed with λ2 gRNA and λ2 ssRNA targets in the presence of PAMmers with 5′-extensions of variable length. The ssRNA cleavage efficiency decreases as the PAMmer extends further into the target region, as indicated by the fraction of RNA cleaved after 1 h. **b**, Binding assays were conducted as described in the Methods, using mostly the same panel of 5′-extended PAMmers as in **a**. Binding reactions contained Cas9 programmed with λ2 gRNA and either λ2 (on-target) or λ3 (off-target) ssRNA in the presence of cognate PAMmers with 5′-extensions of variable length. The binding specificity increases as the PAMmer extends further into the target region, as indicated by the fraction of λ3 (off-target) ssRNA bound at 3 nM Cas9–gRNA. PAMmers with 5′ extensions also cause a slight reduction in the relative binding affinity of λ2 (on-target) ssRNA.
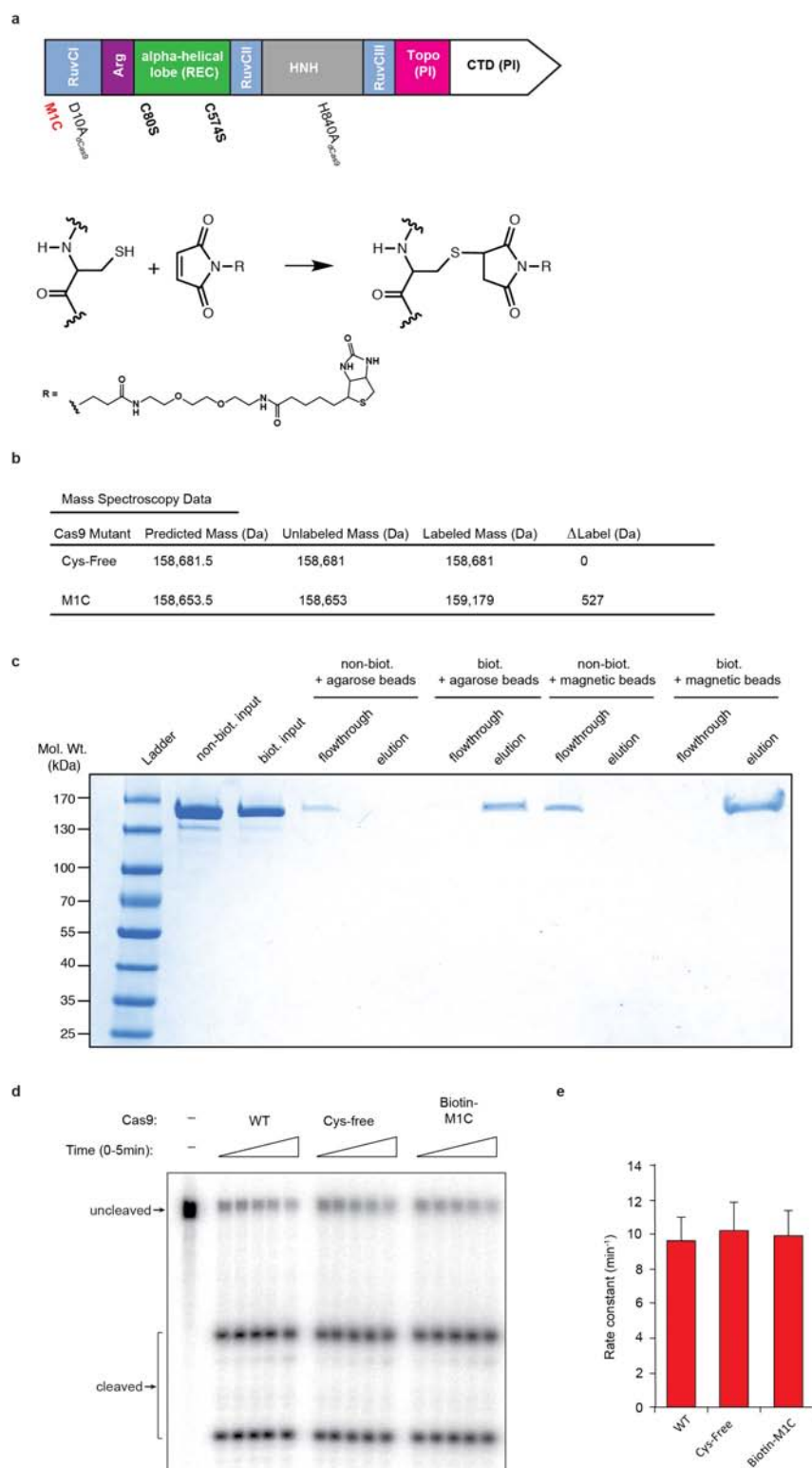
Extended Data Figure 5 | Site-specific biotin labelling of Cas9. a, In order to introduce a single biotin moiety on Cas9, the solvent accessible, non-conserved amino-terminal methionine was mutated to a cysteine (M1C; red text) and the naturally occurring cysteine residues were mutated to serine (C80S and C574S; bold text). This enabled cysteine-specific labelling with EZ-link Maleimide-PEG2-biotin through an irreversible reaction between the reduced sulphydryl group of the cysteine and the maleimide group present on the biotin label. Mutations of dCas9 are also indicated in the domain schematic. **b,** Mass spectrometry analysis of the Cas9 biotin-labelling reaction confirmed that successful biotin labelling only occurs when the M1C mutation is present in the Cys-free background (C80S;C574S). The mass of the Maleimide-PEG2-biotin reagent is 525.6 Da. **c,** Streptavidin bead binding assay with biotinylated (biot.) or non-biotinylated (non-biot.) Cas9 and streptavidin agarose or streptavidin magnetic beads. Cas9 only remains specifically bound to the beads after biotin labelling. **d,** Cleavage assays were conducted as described in the Methods and resolved by denaturing PAGE. Reactions contained 100 nM Cas9 programmed with λ2 gRNA and ~1 nM 5′-radiolabelled λ2 dsDNA target. **e,** Quantified cleavage data from triplicate experiments were fitted with single-exponential decays to calculate the apparent pseudo-first-order cleavage rate constants (average ± standard deviation). Both Cys-free and biotin-labelled Cas9(M1C) retain wild-type activity.
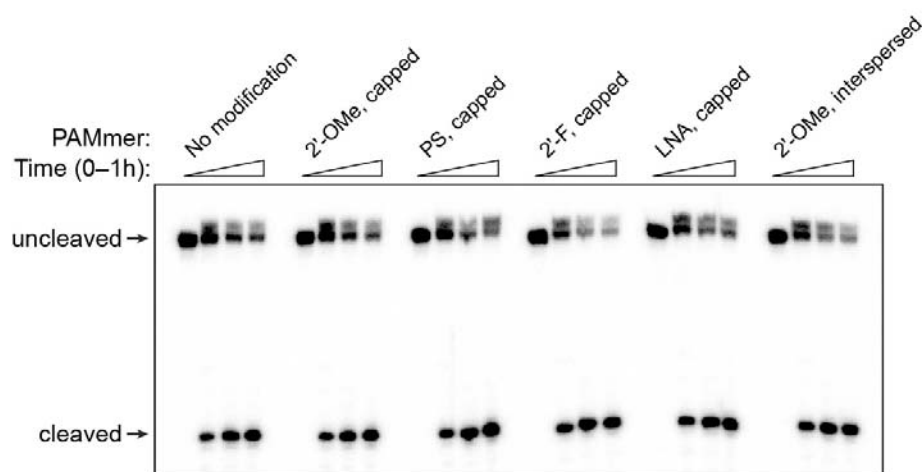
Extended Data Figure 6 | RNA-guided Cas9 can utilize chemically modified PAMmers. Nineteen-nucleotide PAMmer derivatives containing various chemical modifications on the 5′ and 3′ ends (capped) or interspersed throughout the strand still activate Cas9 for cleavage of ssRNA targets. These types of modification are often used to increase the *in vivo* half-life of short oligonucleotides by preventing exo- and endonuclease-mediated degradation. Cleavage assays were conducted as described in the Methods. PS, phosphorothioate bonds; LNA, locked nucleic acid.
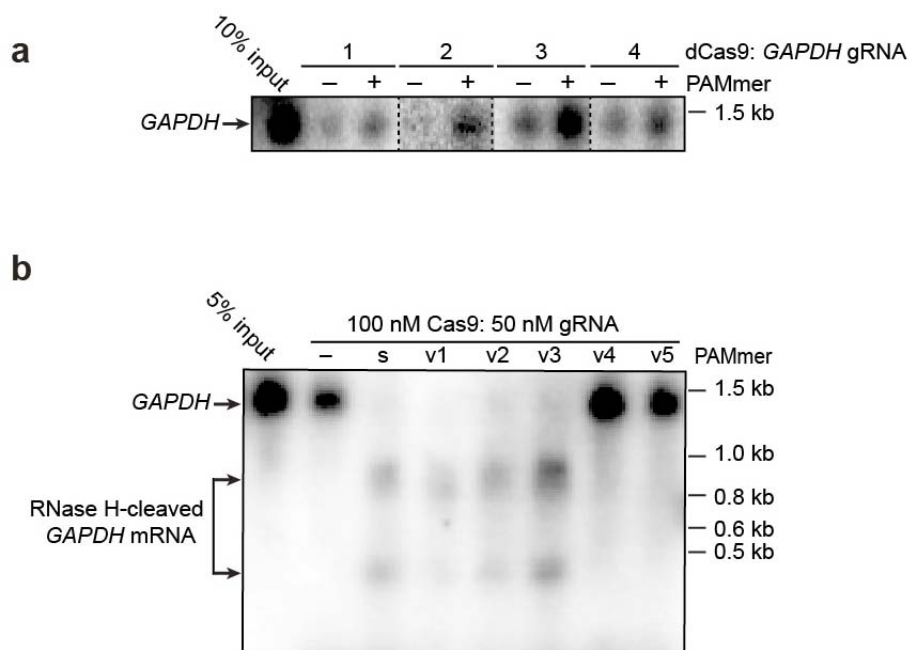
Extended Data Figure 7 | Cas9 programmed with *GAPDH*-specific gRNAs can pull down *GAPDH* mRNA in the absence of PAMmers. a, Northern blot showing that, in some cases, Cas9–gRNA is able to pull down detectable amounts of *GAPDH* mRNA from total RNA without requiring a PAMmer.

b, Northern blot showing that Cas9–gRNA *G*1 is also able to pull down quantitative amounts of *GAPDH* mRNA from HeLa cell lysate without requiring a PAMmer. s, standard; v1-5, increasingly 2′-OMe-modified PAMmers. See Fig. 4g for PAMmer sequences.
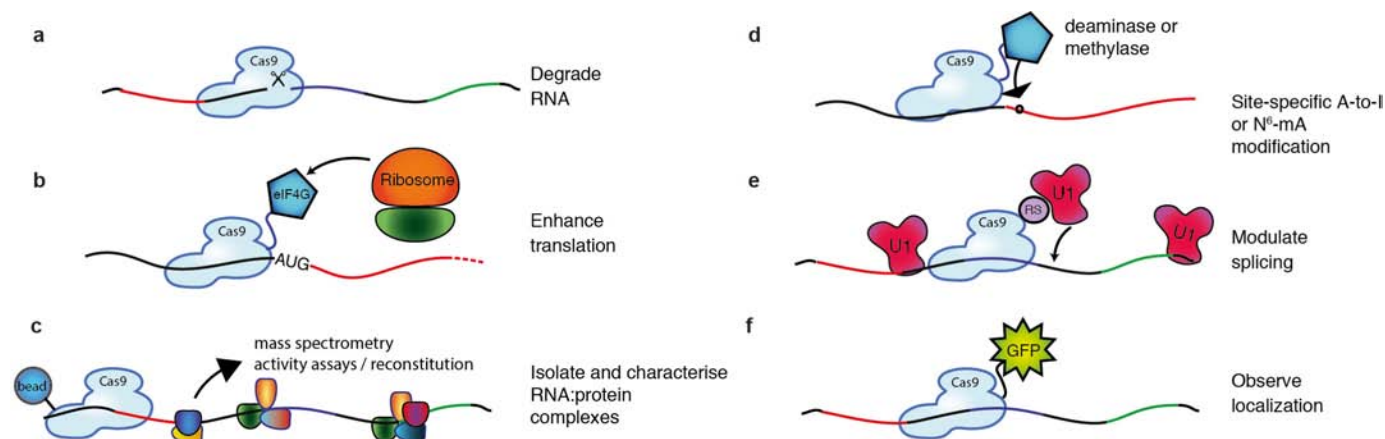
Extended Data Figure 8 | Potential applications of RCas9 for untagged transcript analysis, detection and manipulation. a, Catalytically active RCas9 could be used to target and cleave RNA, particularly those for which RNA-interference-mediated repression/degradation is not possible. **b,** Tethering the eukaryotic initiation factor eIF4G to a catalytically inactive dRCas9 targeted to the 5′ untranslated region of an mRNA could drive translation. **c,** dRCas9 tethered to beads could be used to specifically isolate RNA or native RNA–protein complexes of interest from cells for downstream analysis or assays including identification of bound-protein complexes, probing of RNA structure under native protein-bound conditions, and enrichment of rare transcripts for sequencing analysis. **d,** dRCas9 tethered to RNA deaminase or N6-mA methylase domains could direct site-specific A-to-I editing or methylation of RNA, respectively. **e,** dRCas9 fused to a U1 recruitment domain (arginine- and serine-rich (RS) domain) could be programmed to recognize a splicing enhancer site and thereby promote the inclusion of a targeted exon. **f,** dRCas9 tethered to a fluorescent protein such as GFP could be used to observe RNA localization and transport in living cells. Adapted from Mackay *et al.*[18]

Extended Data Table 1 | λ-Oligonucleotide sequences

Description	Sequence *	Used in
Oligo for T7 promoter, in vitro transcription	5'-**TAATACGACTCACTATA**-3'	NA
λ2-targeting crRNA	5'-GUGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUG-3'	Fig. 1c-e, 3a, 4c-d, ED1-2, 4a
λ3-targeting crRNA	5'-CUGGUGAACUUCCGAUAGUGGUUUUAGAGCUAUGCUGUUUUG-3'	Fig. 3a
λ4-targeting crRNA	5'-CAGAUAUAGCCUGGUGGUUCGUUUUAGAGCUAUGCUGUUUUG-3'	Fig. 3a
ssDNA T7 template†, tracrRNA	5'-AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTGTCCTATAGTGAGTCGTATTA-3'	NA
tracrRNA (nt 15-87)	5'-GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3'	Fig. 1c-e, 3a, 4c-d, ED1-2,4a
λ2-targeting sgRNA T7 template‡	5'-**TAATACGACTCACTATA**GGTGATAAGTGGAATGCCATGGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC**TTTTTTT**-3'	NA
λ2-targeting sgRNA	5'-**GG**UGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC**UUUUUUU**-3'	Fig 2, 3b,d, ED3, 4b
λ2 target dsDNA duplex	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATG**TGG**GCTGTCAAAATTGAGC-3' 3'-CTCACCTTCCTACGGTCACTATTCACCTTACGGTACACCCGACAGTTTTAACTCG-5'	Fig. 1c, 2a, 4c
λ2 ssDNA target strand (used to make heteroduplex DNA:RNA)	3'-CTCACCTTCCTACGGTCACTATTCACCTTACGGTACACCCGACAGTTTTAACTCG-5'	Fig 1c, 2a
λ2 ssDNA non-target strand (used to make heteroduplex DNA:RNA)	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATG**TGG**GCTGTCAAAATTGAGC-3'	Fig 1c, 2a, 3d, ED3
λ2 ssRNA target strand T7 template	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC**CTATAGTGAGTCGTATTA**-3'	NA
λ2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUAUUCACCUUACGGUACACCCGACAGUUUUAACUCG-5'	Fig. 1c-e, 2, 3, 4, ED1-4
λ2 ssRNA non-target strand T7 template	5'-GCTCAATTTTGACAGCCCACATGGCATTCCACTTATCACTGGCATCCTTCCACTC**TATAGTGAGTCGTATTA**-3'	NA
λ2 ssRNA non-target strand (used to make dsRNA)	5'-**GG**AGUGGAAGGAUGCCAGUGAUAAGUGGAAUGCCAUG**TGG**GCUGUCAAAAUUGAGC-3'	Fig. 1c, 2a
19 nt λ2 DNA PAMmer	5'-**TGG**GCTGTCAAAATTGAGC-3'	Fig. 1c-e, 2, 3a-b, ED 1-4
18 nt λ2 'GG' DNA PAMmer	5'-**GG**GCTGTCAAAATTGAGC-3'	Fig. 1c, 2
19 nt λ2 DNA mutated PAMmer	5'-ACCGCTGTCAAAATTGAGC-3'	Fig. 1c, 2c
16 nt λ2 DNA 'PAM-less' PAMmer	5'-GCTGTCAAAATTGAGC-3'	Fig. 1c, 2c
18 nt λ2 RNA PAMmer	5'-**GG**GCUGUCAAAAUUGAGC-3'	Fig. 1c, 2a
5 nt λ2 DNA PAMmer	5'-**TGG**GC-3'	Fig. 1e, 2c
10 nt λ2 DNA PAMmer	5'-**TGG**GCTGTCA-3'	Fig. 1e, 2c
15 nt λ2 DNA PAMmer	5'-**TGG**GCTGTCAAAATT-3'	Fig. 1e, 2c
λ3 ssRNA target strand T7 template	5'-AACGTGCTGCGGCTGGCTGGTGAACTTCCGATAGTGCGGGTGTTGAATGATTTCC**TATAGTGAGTCGTATTA**-3'	NA
λ3 ssRNA target strand	3'-UUGCACGACGCCGACCGACCACUUGAAGGCUAUCACGCCCACAACUUACUAAAGG-5'	Fig. 3a,b,d, ED3, 4b
λ4 ssRNA target strand T7 template	5'-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTCAGGCGGCGCATTTTTATTGCC**TATAGTGAGTCGTATTA**-3'	NA
λ4 ssRNA target strand	3'-AGUGUUGUUACUCACCGUCUAUAUCGGACCACCAAGUCCGCCGCGUAAAAAUAACGG-5'	Fig. 3a,b,d, ED3
λ3 ssDNA non-target strand	5'-AACGTGCTGCGGCTGGCTGGTGAACTTCCGATAGTG**CGG**GTGTTGAATGATTTCC-3'	Fig. 3d, ED3
λ4 ssDNA non-target strand	5'-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTC**AGG**CGGCGCATTTTTATTG-3'	Fig. 3d, ED3
19 nt λ3 DNA PAMmer	5'-**CGG**GTGTTGAATGATTTCC-3'	Fig. 3a,b,d, ED3, 4
19 nt λ4 DNA PAMmer	5'-**AGG**CGGCGCATTTTTATTG-3'	Fig. 3a,b,d, ED3
21 nt λ2 5'-extended DNA PAMmer	5'-TG**TGG**CTGTCAAAATTGAGC-3'	Fig. 4c, ED 4a,b
21 nt λ3 5'-extended DNA PAMmer	5'-TG**CGG**GTGTTGAATGATTTCC-3'	ED 4b
24 nt λ2 5'-extended DNA PAMmer	5'-CCATG**TGG**CTGTCAAAATTGAGC-3'	ED 4a,b
24 nt λ3 5'-extended DNA PAMmer	5'-TAGTG**CGG**GTGTTGAATGATTTCC-3'	ED 4b
27 nt λ2 5'-extended DNA PAMmer	5'-ATGCCATG**TGG**CTGTCAAAATTGAGC-3'	Fig. 4f,g, ED 4a,b
27 nt λ3 5'-extended DNA PAMmer	5'-CGATAGTG**CGG**GTGTTGAATGATTTCC-3'	ED 4b
30 nt λ2 5'-extended DNA PAMmer	5'-GGAATGCCATG**TGG**CTGTCAAAATTGAGC-3'	ED 4a,b
30 nt λ3 5'-extended DNA PAMmer	5'-TTCCGATAGTG**CGG**GTGTTGAATGATTTCC-3'	ED 4b
33 nt λ2 5'-extended DNA PAMmer	5'-AGTGGAATGCCATG**TGG**CTGTCAAAATTGAGC-3'	ED 4a,4b
33 nt λ3 5'-extended DNA PAMmer	5'-AACTTCCGATAGTG**CGG**GTGTTGAATGATTTCC-3'	ED 4b
36 nt λ2 5'-extended DNA PAMmer	5'-ATAAGTGGAATGCCATG**TGG**CTGTCAAAATTGAGC-3'	ED 4a
39 nt λ2 5'-extended DNA PAMmer	5'-GTGATAAGTGGAATGCCATG**TGG**CTGTCAAAATTGAGC-3'	ED 4a,4b
39 nt λ3 5'-extended DNA PAMmer	5'-CTGGTGAACTTCCGATAGTG**CGG**GTGTTGAATGATTTCC-3'	Fig. 4b
non-PAM λ2 dsDNA	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGACGCTGTCGTGTCAAAATTGAGC-3' 3'-CTCACCTTCCTACGGTCACTATTCACCTTACGGTACTGCGACAGTTTTAACTCG-5'	Fig. 4c
non-PAM λ2 ssRNA target strand T7 template	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGACGCTGTCGTGTCAAAATTGAGC**CTATAGTGAGTCGTATTA**-3'	NA
non-PAM λ2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUAUUCACCUUACGGUACTGGCGACAGUUUUAACUCG-5'	Fig. 4c
λ2 2'OMe capped PAMmer§	5'-*U**GG**GCUGUCAAAAUUGAG*C-3'	ED 6
λ2 PS capped PAMmer§	5'-T*GG*CTGTCAAAATTGAG*C-3'	ED 6
λ2 2'F capped PAMmer§	5'-*U**GG**GCUGUCAAAAUUGAG*C-3'	ED 6
λ2 LNA capped PAMmer§	5'-*T**GG**CTGTCAAAATTGAG*C-3'	ED 6
λ2 19 nt 2'OMe interspersed PAMmer§	5'-*U**GG**G*CUGGC*AAAAUU+GAG*C-3'	ED 6

* Guide crRNA sequences and complementary DNA target strand sequences are shown in red. PAM sites (5'-NGG-3') are highlighted in yellow on the non-target strand when adjacent to the target sequence or in the PAMmer.
† The T7 promoter is indicated in bold (or reverse complement of), as well as 5' G or GG included in the ssRNA product by T7 polymerase.
‡ sgRNA template obtained from pIDT, subsequently linearized by AflII for run-off transcription.
§ Positions of modifications depicted with asterisks preceding each modified nucleotide in each case (except for PS linkages which are depicted between bases).
PS, phosphorothioate bond; NA, not applicable; LNA, locked nucleic acid.

Extended Data Table 2 | Oligonucleotides used in the *GAPDH* mRNA pull-down experiment

Description	Sequence *	Used in
GAPDH-targeting sgRNA 1 T7 template†	5′–**TAATACGACTCACTATA**GGGGCAGAGATGATGACCCTGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	NA
GAPDH-targeting sgRNA 1	5′–**GGGG**CAGAGATGATGACCCTGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	Fig. 4f,g, ED 7
GAPDH-targeting sgRNA 2 T7 template†	5′–**TAATACGACTCACTATAGG**CCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	NA
GAPDH-targeting sgRNA 2	5′–**GG**CCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	Fig. 4f, ED 7
GAPDH-targeting sgRNA 3 T7 template†	5′–**TAATACGACTCACTATAGG**CCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	NA
GAPDH-targeting sgRNA 3	5′–**GG**CCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	Fig. 4f, ED 7
GAPDH-targeting sgRNA 4 T7 template†	5′–**GG**ATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	NA
GAPDH-targeting sgRNA 4	5′–**TAATACGACTCACTATAGG**ATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT–3′	Fig. 4f, ED 7
GAPDH PAMmer 1	5′–ATGACCCT**TGG**GGCTCCCCCCTGCAAA–3′	Fig. 4f,g, ED 7
GAPDH PAMmer 2	5′–TGGATGAC**CGG**GGCCAGGGGTGCTAAG–3′	Fig. 4f, ED 7
GAPDH PAMmer 3	5′–TTGGCAGG**TGG**TTCTAGACGGCAGGTC–3′	Fig. 4f, ED 7
GAPDH PAMmer 4	5′–CCCCAGCG**TGG**AAGGTGGAGGAGTGGG–3′	Fig. 4f, ED 7
GAPDH PAMmer 1 2'OMe v1‡	5′–A*UGACC*CT**AGG***GGCTC*CCCCC*UGCAA*A–3′	Fig. 4g, ED 7
GAPDH PAMmer 1 2'OMe v2‡	5′–*ATG*ACC*CU***AGG***GGC*UCC*CCC*CTG*CAA*A–3′	Fig. 4g, ED 7
GAPDH PAMmer 1 2'OMe v3‡	5′–*ATG*ACCC*U**AGG***GGCT*CCCC*CCTG*CAA*A–3′	Fig. 4g, ED 7
GAPDH PAMmer 1 2'OMe v4‡	5′–*AT*GA*CC*CT***AGG***GG*CT*CC*CC*CC*UG*CA*AA–3′	Fig. 4g, ED 7
GAPDH PAMmer 1 2'OMe v5‡	5′–*AT*GA*CC*CT***AG*G**G*GC*TC*CC*CC*CU*GC*AA*A–3′	Fig. 4g, ED 7
GAPDH cDNA primer Fwd	5′–CTCACTGTTCTCTCCCTCCGC–3′	Fig. 4g,f, ED7
GAPDH cDNA primer Rev	5′–AGGGGTCTACATGGCAACTG–3′	Fig. 4g,f, ED7
β-actin cDNA primer Fwd	5′–AGAAAATCTGGCACCACACC–3′	Fig. 4g,f, ED7
β-actin cDNA primer Rev	5′–GGAGTACTTGCGCTCAGGAG–3′	Fig. 4g,f, ED7

* Guide crRNA sequences and complementary DNA target strand sequences are shown in red. PAM sites (5′-NGG-3′) are highlighted in yellow on the non-target strand when adjacent to the target sequence or in the PAMmer.

† The T7 promoter is indicated in bold (or reverse complement of), as well as 5′ G or GG included in the ssRNA product by T7 polymerase. sgRNAs for *GAPDH* were designed according to Chen *et al.*[32]

‡ Positions of 2′-OMe modifications depicted with asterisks preceding each modified nucleotide.

NA, not applicable.

LETTER

Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation

Harihar Basnet[1,2], Xue B. Su[3], Yuliang Tan[1], Jill Meisenhelder[5], Daria Merkurjev[1,4], Kenneth A. Ohgi[1], Tony Hunter[5], Lorraine Pillus[3] & Michael G. Rosenfeld[1]

Post-translational histone modifications have a critical role in regulating transcription, the cell cycle, DNA replication and DNA damage repair[1]. The identification of new histone modifications critical for transcriptional regulation at initiation, elongation or termination is of particular interest. Here we report a new layer of regulation in transcriptional elongation that is conserved from yeast to mammals. This regulation is based on the phosphorylation of a highly conserved tyrosine residue, Tyr 57, in histone H2A and is mediated by the unsuspected tyrosine kinase activity of casein kinase 2 (CK2). Mutation of Tyr 57 in H2A in yeast or inhibition of CK2 activity impairs transcriptional elongation in yeast as well as in mammalian cells. Genome-wide binding analysis reveals that CK2α, the catalytic subunit of CK2, binds across RNA-polymerase-II-transcribed coding genes and active enhancers. Mutation of Tyr 57 causes a loss of H2B mono-ubiquitination as well as H3K4me3 and H3K79me3, histone marks associated with active transcription. Mechanistically, both CK2 inhibition and the H2A(Y57F) mutation enhance H2B deubiquitination activity of the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex, suggesting a critical role of this phosphorylation in coordinating the activity of the SAGA complex during transcription. Together, these results identify a new component of regulation in transcriptional elongation based on CK2-dependent tyrosine phosphorylation of the globular domain of H2A.

To assess potential tyrosine phosphorylation events in H2A, we individually mutated every tyrosine residue in H2A to phenylalanine and expressed the mutants in 293T cells. Mutation of Tyr 39 and Tyr 57 resulted in a decrease in tyrosine phosphorylation compared to the wild-type protein, indicating that these residues might be phosphorylated (Fig. 1a). Mass spectrometry confirmed phosphorylation of these residues in histone extracts from 293T cells (Supplementary Table 1). The Tyr 57 residue, along with neighbouring residues, is conserved from yeast to mammals (Fig. 1b), and is present in all variants of H2A (Extended Data Fig. 1a). In budding yeast, where genetic manipulation of histones is possible, mutation of the corresponding residue to alanine is lethal, suggesting a critical structural and/or functional contribution of this tyrosine residue[2,3]. Analysis of histones in wild-type mononucleosomes and those containing the H2A(Y57F) mutant showed similar stoichiometry, suggesting that the Y57F mutation is unlikely to affect the structural integrity of nucleosomes (Extended Data Fig. 1b). Hence, we tested whether the structurally conservative substitution of tyrosine with phenylalanine would
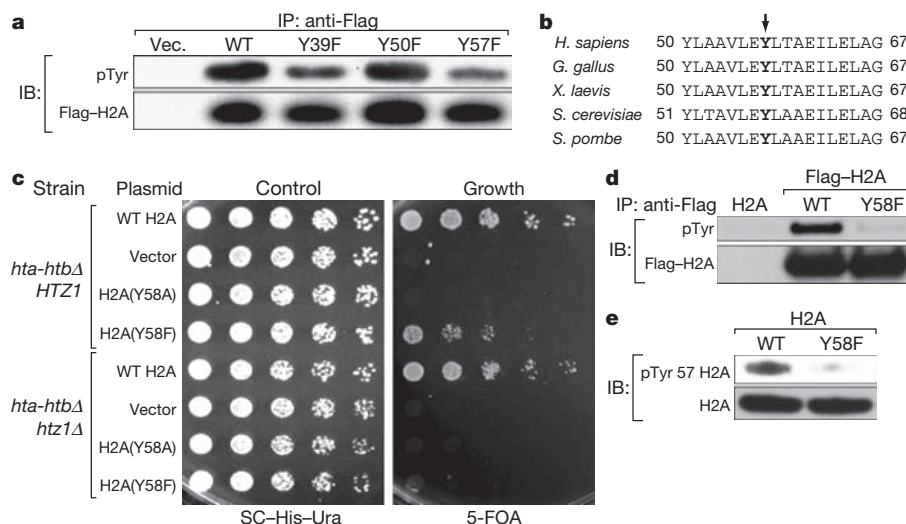
Figure 1 | The conserved Tyr 57 residue in H2A is phosphorylated. a, Tyr 57 in H2A is phosphorylated in 293T cells. Flag-tagged H2A mutants were expressed in 293T cells, immunoprecipitated (IP) under denaturing conditions, and immunoblotted (IB) as indicated. Vec., vector; WT, wild type. **b**, Tyr 57 in H2A is highly conserved. Comparison of H2A sequence surrounding the Tyr 57 residue (arrow) in different organisms. *H.*, *Homo*; *G.*, *Gallus*; *X.*, *Xenopus*; *S.*, *Saccharomyces* (for *cerevisiae*); *S.*, *Schizosaccharomyces* (for *pombe*). **c**, Tyr 58 in H2A is functionally important in yeast. Fivefold serial dilutions of the yeast strains lacking H2A (*hta*) and H2B (*htb*) but containing pJH33 (*HTA1-HTB1 HHF2-HHT2 URA3* CEN) were transformed as indicated, and the transformants were plated on SC (synthetic complete supplement)–His–Ura for growth control and 5-fluoroorotic acid (5-FOA) for the removal of pJH33. **d**, **e**, Tyr 58 in H2A is phosphorylated in *S. cerevisiae*. **d**, Flag-tagged wild-type H2A and H2A(Y58F) were immunoprecipitated under denaturing conditions and immunoblotted as indicated. **e**, Whole-cell extracts from yeast strains were prepared under denaturing conditions and immunoblotted as indicated. Data represent three independent experiments.

[1]Howard Hughes Medical Institute, Department of Medicine, University of California San Diego, La Jolla, California 92093, USA. [2]Biomedical Sciences Graduate Program, School of Medicine, University of California San Diego, La Jolla, California 92093, USA. [3]Division of Biological Sciences, Section of Molecular Biology, UCSD Moores Cancer Center, University of California San Diego, La Jolla, California 92093-0347, USA. [4]Bioinformatics and Systems Biology Program, Department of Bioengineering, University of California San Diego, La Jolla, California 92093, USA. [5]Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA.

produce a different phenotype in yeast. Tyr 58 in yeast H2A corresponds to Tyr 57 in mammalian H2A. The H2A(Y58F) mutant was viable and exhibited a slow growth phenotype (Fig. 1c). Notably, the same mutation proved to be lethal in the *HTZ1* (the gene encoding H2AZ (also known as H2AFZ)) null background, and double mutation of the tyrosine residue in both H2A and H2AZ resulted in an extremely slow growth phenotype (Fig. 1c and Extended Data Fig. 1c). Next, we tested if this site is phosphorylated in yeast. Immunoprecipitated Flag-tagged H2A(Y58F) showed reduced tyrosine phosphorylation compared to the wild-type protein (Fig. 1d), suggesting that this residue is phosphorylated.

To confirm Tyr 57 phosphorylation and investigate its function, an antibody specific for phosphorylated Tyr (pTyr) 57 H2A was developed. This antibody detected proteins corresponding to the size of H2A and ubiquitinated H2A in 293T cells (Extended Data Fig. 1d). Peptide blocking assays and dot blot assays verified the specificity of the antibody and treatment with calf intestinal phosphatase further validated its phospho-specificity (Extended Data Fig. 1d–f). Use of this antibody confirmed Tyr 58 phosphorylation in yeast H2A (Fig. 1e). Collectively, these results demonstrate that Tyr 57 in H2A is phosphorylated and that this phosphorylation is conserved from yeast to mammals.

To identify the kinase(s) that mediate(s) phosphorylation of Tyr 57 in H2A in mammals, we performed mass spectrometry analysis of proteins interacting with H2A in 293T cells. To be consistent with yeast, we used H2AX (also known as H2AFX), a common H2A variant that has closer sequence homology to yeast H2A, for co-immunoprecipitation and *in vitro* kinase assays. Mass spectrometry data revealed that the CK2α catalytic subunit of CK2 interacts preferentially with H2A(Y57F) compared to wild-type H2A (Supplementary Table 2). This interaction was further

verified by immunoblotting, which revealed a higher level of CK2α associated with H2A(Y57F) (Fig. 2a). One implication of this interaction is that CK2 may phosphorylate Tyr 57 in H2A, and the Y57F mutation stabilizes the enzyme–substrate interaction, analogous to substrate trapping approaches that have been successfully used to identify substrates of tyrosine phosphatases[4]. Although CK2 is considered primarily to be a Ser/Thr kinase, two studies have reported its tyrosine phosphorylation activity, thus implicating it as a dual-specificity kinase[5,6]. To investigate the potential roles of CK2 in H2A Tyr 57 phosphorylation, we tested whether CK2α phosphorylates the tyrosine residue (Tyr 57) in H2A in an *in vitro* assay using full-length H2A or nucleosomes. The kinase assay revealed that CK2α phosphorylates Tyr 57 in H2A, acting preferentially in the context of nucleosomes (Fig. 2b). This phosphorylation was inhibited by tetrabromobenzimidazole (TBBz), a chemical inhibitor of CK2[7]. To further establish the tyrosine kinase activity of CK2α, we performed phosphoamino acid analysis (PAA) of phosphorylated H2A from nucleosomes using [γ-^{32}P]ATP. We found that CK2α phosphorylates tyrosine as well as serine residues in H2A, but does not phosphorylate threonine residues, and that Tyr 57 is a phosphorylation site, as demonstrated by the reduced tyrosine phosphorylation in H2A(Y57F) compared to wild-type H2A (Extended Data Fig. 2a).

Next we investigated whether CK2 is necessary for H2A Tyr 57 phosphorylation *in vivo*. CK2α knockdown in 293T cells reduced the level of Tyr 57 phosphorylation in H2A (Fig. 2c), supporting an *in vivo* role of CK2α in regulating this phosphorylation. Moreover, a dose-dependent decrease in H2A Tyr 57 phosphorylation was observed upon treatment with TBBz (Fig. 2d), further supporting the role of CK2 in Tyr 57 phosphorylation in H2A. Together, these results provide strong evidence for a function of CK2 in H2A Tyr 57 phosphorylation.

To investigate the physiological significance of Tyr 57 phosphorylation in H2A, we examined in yeast the impact of the H2A(Y58F) mutation on other important histone marks. We found that the H2A(Y58F) mutation resulted in a loss of H2B mono-ubiquitination, as well as trimethylation of H3K4 and H3K79 (Fig. 3a). H3K27 acetylation showed a modest increase, and all other histone modifications tested were unaffected (Extended Data Fig. 3a). Notably, the Y57F mutation also lowered the level of H2A mono-ubiquitination in 293T cells (Extended Data Fig. 3b, c). The role of H2A Tyr 58 phosphorylation as a potential regulator of H2B mono-ubiquitination is of particular significance because this modification has an established role in transcriptional elongation[8,9], thus potentially linking H2A Tyr 58 phosphorylation to transcriptional elongation. Further, yeast with the H2A(Y58F) mutation exhibited increased sensitivity to 6-azauracil (Extended Data Fig. 3d), indicating a defect in transcriptional elongation[10].

To assess the role of Tyr 58 phosphorylation in transcriptional elongation, binding of RNA polymerase II (Pol II) in actively transcribed genes was evaluated by chromatin immunoprecipitation (ChIP) followed by quantitative real-time polymerase chain reaction (qPCR). Pol II binding was reduced in the gene body of a housekeeping gene, *PYK1*, as well as a number of heat-shock-induced genes[11] upon heat shock in the H2A(Y58F) mutant (Fig. 3b). Pol II binding was further reduced in H2A(Y58F) and H2AZ(Y65F) mutants (Fig. 3b). The decrease in Pol II binding in the H2A(Y58F) mutant was not due to reduced Pol II expression (Extended Data Fig. 3e). Consistent with the defect in transcriptional elongation, a decreased level of messenger RNA of the corresponding genes was observed in H2A(Y58F) mutants (Extended Data Fig. 3f), whereas H2AZ(Y65F) mutation alone caused only a mild defect in transcription of heat-shock-induced genes (Extended Data Fig. 3g). Furthermore, in 293T cells, H2A Tyr 57 phosphorylation, like H2B mono-ubiquitination, was correlated with transcriptional elongation events, as demonstrated when transcriptional elongation was blocked with flavopiridol treatment[12], and induced by washing out the drug (Extended Data Fig. 3h). H3K4me2, a control histone mark, did not change in this assay (Extended Data Fig. 3h). Collectively, these results indicate a conserved role of Tyr 57/58 phosphorylation in H2A in regulating transcriptional elongation.
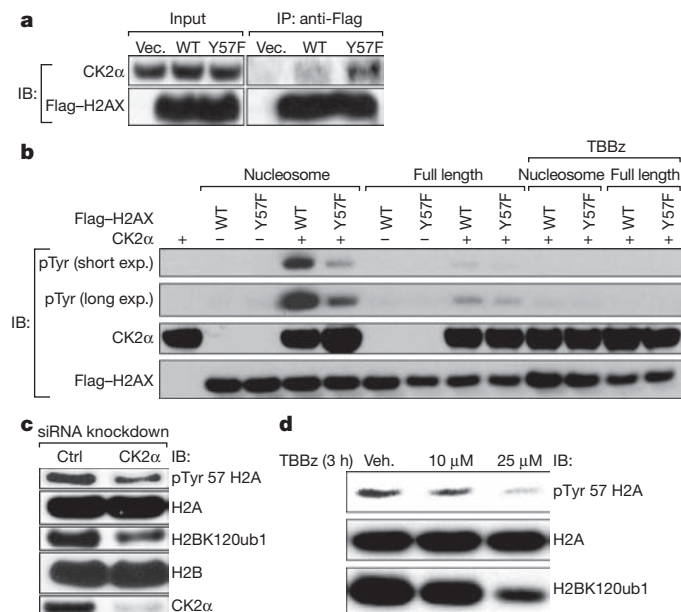
Figure 2 | CK2 phosphorylates Tyr 57 in H2A. a, CK2α interacts preferentially with the H2A(Y57F) mutant. Flag-tagged wild-type (WT) H2AX and H2AX(Y57F) were expressed in 293T cells, immunoprecipitated (IP) and immunoblotted (IB). **b**, CK2 phosphorylates Tyr 57 in H2A in nucleosomes *in vitro*. *In vitro* kinase assays were performed using recombinant glutathione *S*-transferase (GST)-CK2α, and full-length or nucleosomal Flag-tagged H2AX purified from 293T cells, and were immunoblotted to examine tyrosine phosphorylation. **c, d**, CK2 phosphorylates Tyr 57 in H2A *in vivo*. **c**, CK2α was knocked down in 293T cells, and nuclear extracts were immunoblotted. Ctrl indicates scrambled short interfering RNA (siRNA). **d**, Nuclear extract from 293T cells treated with vehicle (Veh.) or TBBz for 3 h were immunoblotted. Data represent three independent experiments.
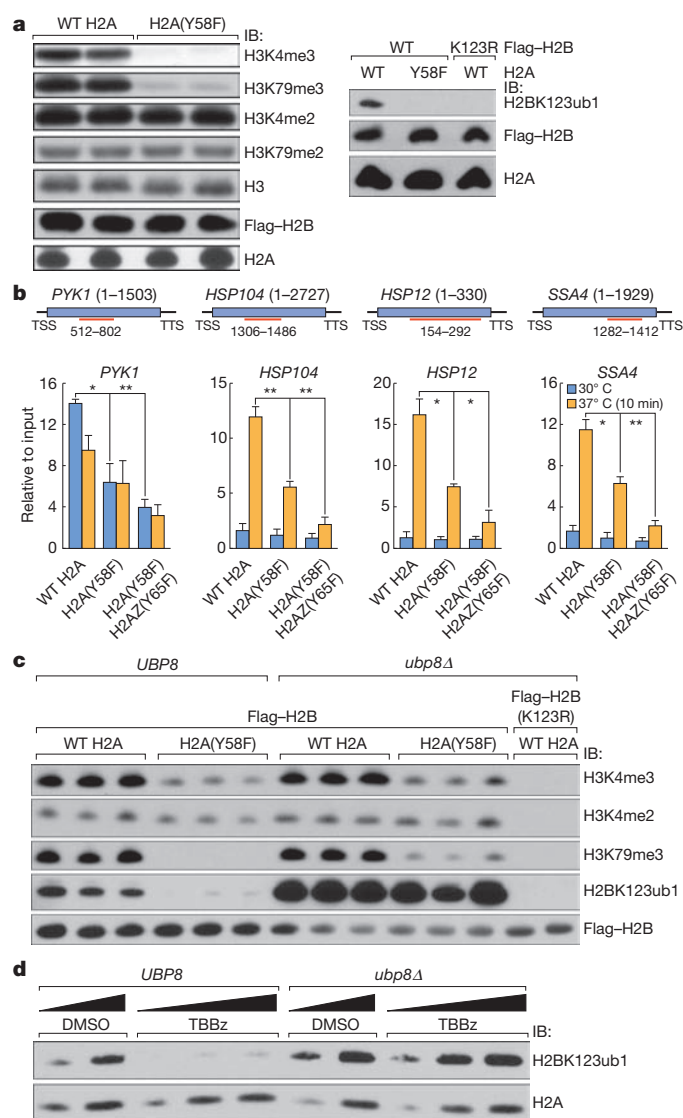
Figure 3 | The H2A(Y58F) mutation enhances H2B deubiquitination, and impairs transcriptional elongation in yeast. a, The H2A(Y58F) mutation affects several histone modifications. Whole-cell extracts from indicated strains were immunoblotted (IB). **b**, H2A(Y58F) mutation impairs transcriptional elongation. Pol II binding in the indicated genes was measured by ChIP-qPCR in yeast strains grown at 30 °C or at 37 °C for 10 min (($n = 3$, mean $\pm$ s.e.m.), $*P < 0.05$, $**P < 0.01$). The ORF of the genes and the regions amplified by the primer pairs are shown. P values were calculated by two-tailed Student's t-tests. TSS, transcription start site; TTS, transcription termination site. **c**, *UBP8* deletion rescues the defect in H2B mono-ubiquitination in the H2A(Y58F) mutant yeast. Whole-cell extracts from indicated yeast strains were immunoblotted. **d**, CK2 prevents the deubiquitination of H2B. Wild-type (*UBP8*) or *ubp8Δ* cells were treated with vehicle (dimethylsulphoxide, DMSO) or TBBz (25 μM) for 3 h and whole-cell extracts were immunoblotted. Data represent three independent experiments.

To address the mechanism through which H2A Tyr 57/58 phosphorylation regulates H2B mono-ubiquitination, we examined the recruitment of proteins known to be involved in establishing H2B mono-ubiquitination, such as Paf1, Rtf1 and Rad6 (refs 13–15), by ChIP-qPCR. The binding of Paf1 and Rtf1 was comparable in wild-type and H2A(Y58F) yeast strains, whereas Rad6 binding was slightly reduced in the genes tested in the H2A(Y58F) mutant strain (Extended Data Fig. 4a–c). The effect of the H2A(Y58F) mutation was further evaluated in yeast lacking *UBP8*, which encodes a major H2B deubiquitinase that is a component of the SAGA complex. Deletion of *UBP8* restored H2B mono-ubiquitination

in the H2A(Y58F) mutant to the wild-type level (Fig. 3c), suggesting that the defect in H2B mono-ubiquitination in the H2A(Y58F) mutant occurs through Ubp8 deubiquitinase activity, and not through defective ubiquitination machinery. Likewise, CK2 inhibition reduced H2B mono-ubiquitination in wild-type yeast while having no effect in *UBP8* mutants (Fig. 3d), further supporting the role of CK2-mediated H2A Tyr 58 phosphorylation in preventing H2B deubiquitination. Notably, despite the complete rescue of H2B mono-ubiquitination, *UBP8* deletion only partially rescued H3K79me3, and the level of H3K4me3 remained low in the H2A(Y58F) mutant (Fig. 3c). Furthermore, deletion of *UBP8* did not rescue defects in Pol II binding, the transcript levels, or the slow growth phenotype of the H2A(Y58F) mutant (Extended Data Fig. 4d–f). These results suggest that the physiological effects of Tyr 58 mutation in H2A are linked to, but extend beyond, the loss of H2B mono-ubiquitination.

Next, we investigated whether knockdown or inhibition of CK2 phenocopies the effects of H2A(Y58F) mutation in the regulation of transcriptional elongation. Consistent with the result in yeast, H2B mono-ubiquitination was reduced upon CK2α knockdown or inhibition of CK2 kinase activity in 293T cells (Fig. 2c, d). Likewise, Pol II binding in active genes in LNCaP human prostate carcinoma cells was impaired in gene bodies but not in the promoter regions upon CK2 inhibition, as determined by ChIP followed by sequencing (ChIP-seq) (Fig. 4a). A travelling ratio plot of Pol II[16,17] showed a significant shift in CK2-inhibited cells (Fig. 4a), suggesting that CK2 kinase activity is required for transcriptional elongation in gene bodies. In agreement with this, dihydrotestosterone (DHT)-induced transcriptional activation of androgen-receptor-regulated genes was impaired in LNCaP cells treated with TBBz (Extended Data Fig. 5a).

To understand the molecular aspects of the role of CK2 in transcriptional elongation, genome-wide localization of CK2α in LNCaP cells was determined by ChIP-seq. Like Pol II, CK2α showed binding to actively transcribed genes across gene bodies, although its binding profile was distinct from that of Pol II (Fig. 4b). Meta-analysis of the top 10% of active genes, based on global run-on sequencing results in LNCaP cells[18], revealed that CK2α globally co-localizes with Pol II (Fig. 4c). Consistent with the genome-wide binding pattern, CK2α immunoprecipitated with the phosphorylated carboxy-terminal domain (CTD) of Pol II, which is localized in the promoters (pSer 5) and gene bodies (pSer 2) of active genes (Extended Data Fig. 5b). We also found CK2α binding in intergenic regions that co-localized with H3K4me1 and H3K27ac marks, histone modifications that co-localize with active enhancers[19], and LNCaP cell-type-specific androgen receptor enhancers (Fig. 4d and Extended Data Fig. 5c), suggesting that the intergenic CK2α peaks are in enhancer regions. Inhibition of CK2 also caused stalling of Pol II in the androgen-receptor-bound enhancers (Fig. 4e, Extended Data Fig. 5d), underpinning the function of CK2 in transcriptional elongation in both gene bodies and enhancers.

We asked if CK2 also regulates transcriptional elongation in yeast, and if so, whether H2A Tyr 58 phosphorylation is a key player, among the many other substrates of CK2[20]. Inhibition of CK2 kinase activity resulted in a decrease in the recruitment of Pol II in both the promoter region as well as gene bodies of the tested genes in wild-type yeast, but did not have additive effects in the H2A(Y58F) yeast (Fig. 4f). It is noteworthy that both CK2 inhibition and H2A(Y58F) mutation did not result in promoter-proximal pausing in the genes tested, consistent with the observation that regulation by promoter-proximal pausing is rare in yeast[21]. Collectively, these results demonstrate that CK2 has a deeply conserved role in transcriptional elongation both in gene bodies and enhancers, and that H2A Tyr 58 phosphorylation is critical in this regulation.

This study identifies a new H2A modification, phosphorylation of Tyr 57/58, which provides new insight into how two important protein complexes, SAGA and Paf1, with opposite enzymatic effects on H2B mono-ubiquitination, might be coordinated during transcriptional elongation. The data further emphasize the key significance of this delicate coordination as demonstrated by defects in transcriptional elongation upon mutation of the conserved phosphorylation site. A moderate increase
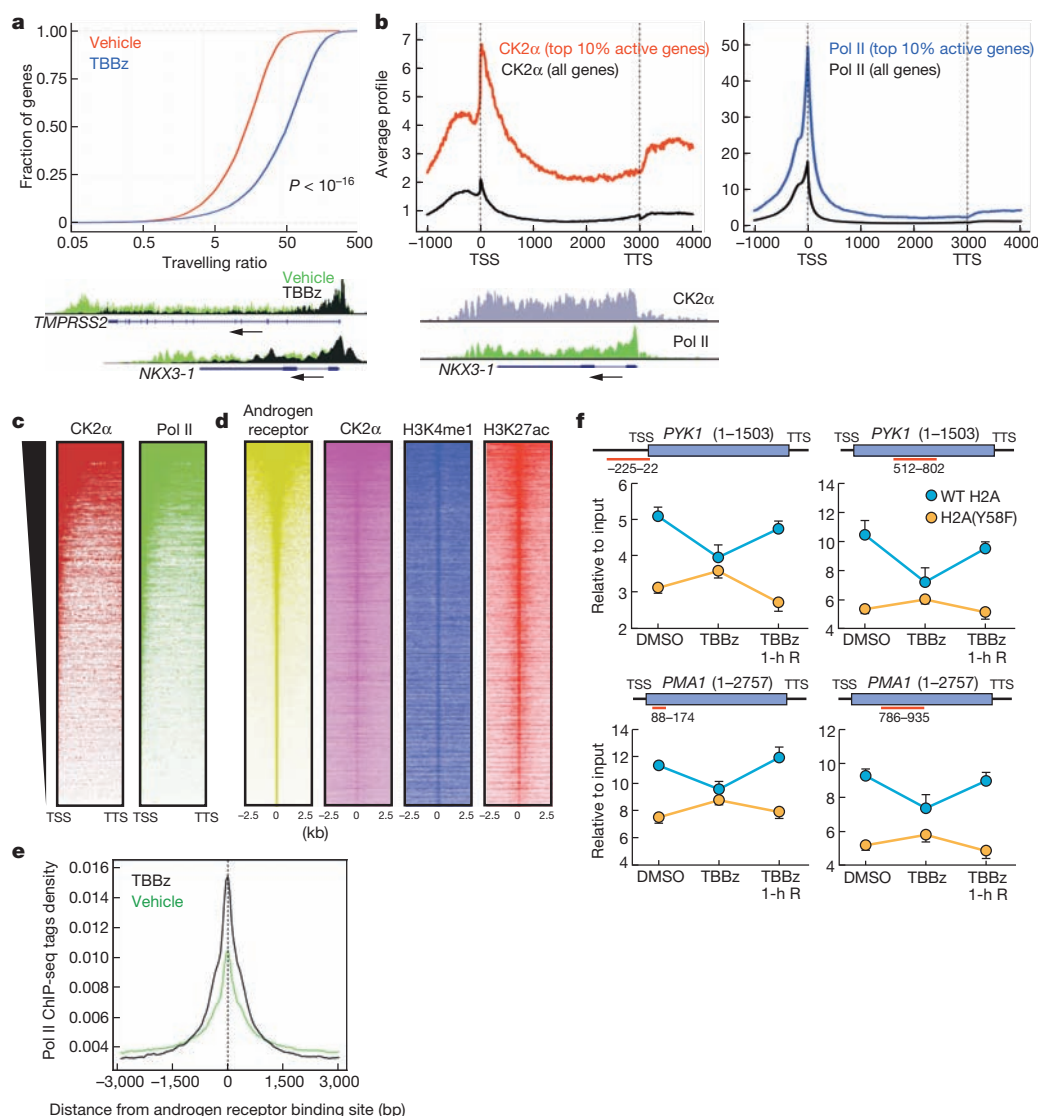
Figure 4 | CK2 regulates transcriptional elongation. a, CK2 kinase activity is required for promoter-proximal pause release in mammalian cells. The Pol II travelling ratio was plotted for LNCaP cells treated with vehicle or TBBz for 2.5 h ($P < 10^{-16}$). P value was calculated using a Kolmogorov–Smirnov test. Overlay of Pol II occupancy in representative genes (*TMPRSS2* and *NKX3-1*) are shown. **b, c**, CK2α binds across the actively transcribed genes globally. **b**, CK2α and Pol II occupancy were determined in the top 10% active genes ($n = 3,162$) and all genes by CEAS[24]. The length of all gene bodies is normalized to 3 kilobases (kb). Enrichment of CK2α and Pol II at a representative active gene (*NKX3-1*) is shown. **c**, Heat map of Pol II and CK2α binding profile in top 10% active genes from transcription start site to transcription termination site is shown. **d**, CK2α binds to active enhancers. Heat maps of androgen receptor

CK2α, H3K4me1 (ref. 18), and H3K27ac (ref. 18) signals over androgen-receptor-enriched regions[18] in LNCaP cells were determined. **e**, CK2 regulates transcriptional elongation in enhancers. Shown are Pol II ChIP tag density plots centred at androgen-receptor-enriched enhancers. bp, base pairs. **f**, H2A Tyr 58 phosphorylation is critical for CK2-mediated regulation of transcriptional elongation in yeast. Cells expressing either wild-type H2A or H2A(Y58F) mutant were treated with vehicle (DMSO, dimethylsulphoxide) or TBBz (25 μM) for 3 h, or treated with TBBz for 3 h followed by 1-h inhibition release (TBBz 1-h R), and Pol II binding in the indicated genes was measured by ChIP-qPCR ($n = 3$, mean ± s.e.m.). The ORF of the genes and the regions amplified by the primer pairs are shown. Data represent two independent experiments for ChIP-seq (**a**–**e**) and three for ChIP-qPCR (**f**).

in Gcn5–SAGA-mediated H3K27 acetylation[22] in the H2A(Y58F) mutant yeast suggests that phosphorylation may antagonize multiple activities of SAGA. Such antagonism could explain the partial rescue of the defects in the H2A(Y58F) mutant upon deletion of *UBP8*, as the other modules of SAGA remain functional in the absence of Ubp8[23]. Although unlikely, the potential role in transcriptional elongation of the hydroxyl group of Tyr 58, rather than phosphorylation itself, cannot yet be dismissed. Assays using synthetic nucleosomes with constitutively phosphorylated H2A Tyr 57/58 may ultimately be used to further define the role of this site in transcriptional elongation. This study also emphasizes the functional significance of the tyrosine kinase activity of CK2, and encourages the search for other tyrosine substrates of CK2. Importantly, the identification of a highly conserved role of CK2 in regulating transcriptional elongation

in both gene bodies and enhancer regions adds yet another layer to understanding transcription.

1. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
2. Nakanishi, S. *et al.* A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nature Struct. Mol. Biol.* **15**, 881–888 (2008).
3. Matsubara, K., Sano, N., Umehara, T. & Horikoshi, M. Global analysis of functional surfaces of core histones with comprehensive point mutants. *Genes Cells* **12**, 13–33 (2007).

4. Flint, A. J., Tiganis, T., Barford, D. & Tonks, N. K. Development of 'substrate-trapping' mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl Acad. Sci. USA* **94,** 1680–1685 (1997).
5. Wilson, L. K., Dhillon, N., Thorner, J. & Martin, G. S. Casein kinase II catalyzes tyrosine phosphorylation of the yeast nucleolar immunophilin Fpr3. *J. Biol. Chem.* **272,** 12961–12967 (1997).
6. Vilk, G. *et al.* Protein kinase CK2 catalyzes tyrosine phosphorylation in mammalian cells. *Cell. Signal.* **20,** 1942–1951 (2008).
7. Pagano, M. A. *et al.* 2-Dimethylamino-4,5,6,7-tetrabromo-1*H*-benzimidazole: a novel powerful and selective inhibitor of protein kinase CK2. *Biochem. Biophys. Res. Commun.* **321,** 1040–1044 (2004).
8. Xiao, T. *et al.* Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol. Cell. Biol.* **25,** 637–651 (2005).
9. Tanny, J. C., Erdjument-Bromage, H., Tempst, P. & Allis, C. D. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev.* **21,** 835–847 (2007).
10. Uptain, S. M., Kane, C. M. & Chamberlin, M. J. Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66,** 117–172 (1997).
11. Ferguson, S. B. *et al.* Protein kinase A regulates constitutive expression of small heat-shock genes in an Msn2/4p-independent and Hsf1p-dependent manner in *Saccharomyces cerevisiae. Genetics* **169,** 1203–1214 (2005).
12. Chao, S.-H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA Polymerase II transcription *in vivo. J. Biol. Chem.* **276,** 31793–31799 (2001).
13. Robzyk, K., Recht, J. & Osley, M. A. Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287,** 501–504 (2000).
14. Wood, A., Schneider, J., Dover, J., Johnston, M. & Shilatifard, A. The Paf1 complex is essential for histone monoubiquitination by the Rad6–Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J. Biol. Chem.* **278,** 34739–34742 (2003).
15. Ng, H. H., Dole, S. & Struhl, K. The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *J. Biol. Chem.* **278,** 33625–33628 (2003).
16. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).
17. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell* **24,** 747–757 (2006).
18. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394 (2011).
19. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell type-specific gene expression. *Nature* **459,** 108–112 (2009).
20. Meggio, F. & Pinna, L. A. One-thousand-and-one substrates of protein kinase CK2? *FASEB J.* **17,** 349–368 (2003).
21. Wade, J. T. & Struhl, K. The transition from transcriptional initiation to elongation. *Curr. Opin. Genet. Dev.* **18,** 130–136 (2008).
22. Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol. Cell* **8,** 473–479 (2001).
23. Henry, K. W. *et al.* Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev.* **17,** 2648–2663 (2003).
24. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25,** 2605–2606 (2009).

METHODS

Cell culture, short interfering RNA, primers, plasmids, transfection, antibodies and kinase inhibitors. LNCaP and 293T cells were cultured in F12 medium supplemented with 10% fetal bovine serum (FBS) and glutamine. For the DHT-treatment experiments, LNCaP cells were cultured in deficient DME high glucose medium with 5% FBS (charcoal dextran filtered) for 3–4 days. Short interfering RNA (siRNA) against CK2α was from Santa Cruz (sc-29918). Cells were transfected using lipofectamine 2000 (Invitrogen) using the manufacturer's protocol. Mutagenesis was done using Quikchange Lightening Mutagenesis Kit following the manufacturer's recommended protocol. The following antibodies were used in this study: pTyr 57 H2A antibody was generated by Biomatik Company using pTyr 57 H2A peptide as an antigen (LE(pY)LTAEILELAGNC), purified, and positively selected using a pTyr 57 H2A peptide column, and negatively selected using a Tyr 57 H2A peptide column; anti-H2A (Abcam no. ab18255), anti-CK2α (Abcam no. ab 70774), anti-H2BK120ub1 (Cell Signaling no. 5546S), anti-Flag (Sigma M2), anti-RNA polymerase II (Santa Cruz N-20 (for mammalian cells), Abcam no. ab817 (for yeast cells)), anti-pSer 2 Pol II (Abcam no. ab5095), anti-pSer 5 Pol II (Abcam no. ab5131), anti-pTyr (Millipore no. 05-321 (4G10)), anti-H2A (yeast) (Active Motif no. 39236), anti-H4K4me3 (Active Motif no. 39159), anti-H3K4me2 (Millipore no. 07-030), anti-H3K4me1 (Millipore no. 07-436), anti-pSer 10-H3 (Millipore no. 06-570), anti-H3K27ac (Abcam no. ab4729), anti-H3K79me2 (Active Motif no. 39924), anti-H3K36me2 (Active Motif no. 39255) and anti-H2AK119ub (Cell Signaling no. 8240S). TBBz and flavopiridol were from Sigma. Primers used in this study are listed in Supplementary Table 3.

Chromatin immunoprecipitation. Cells were grown to 90–95% confluence, fixed with 1% formaldehyde for 15 min for Pol II ChIP, and with di-succinimidyl glutarate (DSG) for 45 min followed by 10 min fixation with 1% formaldehyde for CK2α ChIP. Fixations were performed at room temperature. To terminate crosslinking, fixed cells were incubated with glycine (1.25 mM) for 10 min. Nuclei were prepared as described[25], which were then sonicated in lysis buffer (150 mM NaCl, 1% Triton X-100, 20 mM Tris pH 8.0, 0.1% SDS) using a Bioruptor to fragment chromatin to less than 500 base pairs (bp). Chromatin was pre-cleared with Protein-G magnetic beads, and then immunoprecipitated using 5 μg of antibody per 10-cm plate for each sample. The chromatin antibody mix was incubated with 35 μl protein G-conjugated Dyna beads for 4 h at 4 °C, washed three times with wash buffer (1% Triton X-100, 50 mM Tris pH 8.0, 10% glycerol) with increasing concentrations of NaCl (150 mM, 300 mM and 400 mM), and two times with Tris-EDTA buffer. DNA was eluted in 1% SDS in Tris-EDTA buffer for 45 min at 37 °C, crosslinking was reversed overnight at 65 °C, and DNA was purified using Qiagen columns. Yeast ChIP was performed using 300–400 μg DNA and 2 μg antibody using the protocol as described[26] with some modifications. In brief, cells were fixed for 15 min, and then incubated with glycine at a final concentration of 2.5 mM for 5 min, and cells were lysed using glass beads (5 × 5 min beating with 2 min on ice intervals). Cells were then sonicated to obtain chromatin fragments of less than 1 kb, and the rest of the protocol was similar to the mammalian cells. All the buffers used included fresh complete protease inhibitors (Roche), 1 mM PMSF, 2 mM Na₃VO₄, 10 mM β-glycerol phosphate and 10 mM NaF.

Statistical analysis. P values for ChIP-qPCR and reverse transcription (RT)–PCR were calculated by two-tailed Student's t-tests, type two using Microsoft Excel. The statistical significance of the change of travelling ratio between control and CK2-inhibitor-treated samples was determined using two-tailed Kolmogorov–Smirnov test. In yeast experiments, independent transformants that were processed separately were considered biological replicates, and in mammalian experiments, cells cultured in different plates, and processed separately were considered biological replicates. In the analyses, n represents the number of biological replicates. All independent experiments are biological replicates.

Identification of ChIP-seq peaks. ChIP-seq peaks were identified using HOMER (http://sdcsb.ucsd.edu/resources/homer/). A 200-bp sliding window was used for transcription factors and a 500-bp sliding window was used for histone modifications with the requirement that two peaks are at least 500 bp apart for transcription factors, and 1,250 bp for histone modifications to avoid redundant peak identification. Tag density was calculated by using HOMER and average signal profiles surrounding androgen-receptor-enriched regions were generated with CEAS[24] (cis-regulatory element annotation system) which were visualized with Java TreeView (http://jtreeview.sourceforge.net).

ChIP-seq alignment. DNA was ligated to specific adaptors followed by high-throughput sequencing on Illumina's HiSeq 2000 system according to the manufacturer's instructions. The first 48 bp for each sequence tag returned was aligned to the hg18 (human) assembly using Bowtie2. The data were visualized by preparing custom tracks on the UCSC genome browser by using HOMER. The total number of mappable reads was normalized to 10^7 for each experiment presented.

ChIP-seq data deposition. ChIP-seq data has been deposited in the Gene Expression Omnibus database under accession number GSE58607. Other published sequencing data used in the study are described in ref. 18.

Travelling ratio calculation. The Pol II travelling ratio was defined as the relative ratio of Pol II density in the promoter-proximal region and the gene body. The promoter proximal region refers to the window from −50 bp to +300 bp surrounding the TSS, and the gene body refers to regions from 300 bp downstream of the TSS to 13 kb from the TSS for genes longer than 13 kb, or to the transcription termination site for genes shorter than 13 kb. The significance of the change of the travelling ratio between control and CK2-inhibitor-treated samples was determined by two-tailed Kolmogorov–Smirnov test.

In vitro kinase assay and phosphoamino acid analysis (PAA). In vitro kinase reactions were performed with 100 ng of recombinant GST-tagged CK2α (expressed in E. coli with 0.2 mM isopropyl-β-D-thiogalactoside induction for 3 h at 30 °C) in 1× kinase buffer (20 mM Tris-HCl, 50 mM KCl, 10 mM MgCl₂ pH 7.5) with the addition of 0.2 mM ATP for cold reactions (and 10 μM ATP mixed with 10 μCi of [γ-³²P]ATP for the radioactive reactions). The substrates (Flag-tagged wild-type H2AX and H2AX(Y57F)) were purified from 293T cells expressing Flag-tagged H2AX constructs. For the full-length proteins, histone extracts were immunoprecipitated using Flag antibody, then washed several times with wash buffer (1% Triton X-100, 900 mM NaCl, 20 mM Tris 8.0), treated with calf intestinal phosphatase for 30 min at 37 °C, washed a few more times with wash buffer, and eluted with 3× Flag peptides. Mononucleosomes were prepared as described[25] with minor changes in micrococcal nuclease digestion. In brief, nuclei were isolated from 15-cm fully confluent plates, and DNA was digested in 1.2 ml total volume with 2.5 μl micrococcal nuclease (NEB no. M0247S) for 10 min at 37 °C, and the reaction was stopped by adding 5 mM EGTA, and mononucleosomes were collected by centrifugation. Mononucleosomes were immunoprecipitated with Flag antibody, washed four times with buffer A (340 mM sucrose, 10 mM HEPES pH 7.5, 10% glycerol, 1.5 mM MgCl2, 10 mM KCl) followed by three washes with kinase reaction buffer, then treatment with 500 μM FSBA (Sigma no. F9128-25MG) for 25 min at 37 °C to irreversibly inhibit any potential kinases interacting with the nucleosome. Samples were washed three times with kinase buffer, treated with calf intestinal phosphatase for 30 min at 37 °C, and washed a further three times with buffer A. The bound nucleosomes were eluted with 3× Flag peptides in buffer A. The kinase reactions were carried out for 1 h at 30 °C. For PAA, the samples were separated by SDS–PAGE, transferred to polyvinylidene difluoride membrane, and the membrane corresponding to the mobility of phosphorylated H2AX was excised, and PAA using two-dimensional electrophoresis on thin layer cellulose plates was performed as described[27].

Whole-cell extracts, immunoprecipitation and cell fractionation. Yeast whole-cell extracts were prepared either by breaking the cells with glass beads in PBS or boiling the cells in denaturing buffer (2% SDS with 30 mM dithiothreitol) for 10 min. To immunoprecipitate the Flag-tagged proteins under denaturing conditions, whole-cell extracts were prepared as noted above in denaturing buffer, and the SDS concentration was adjusted to 0.1% by adding dilution buffer (150 mM NaCl, 1% Triton X-100, 20 mM Tris pH 8.0.), then immunoprecipitated overnight using anti-Flag (M2-Sigma) conjugated to agarose beads, and washed five times with dilution buffer. Bound proteins were eluted with 100 μg ml⁻¹ 3× Flag peptides in Tris-buffered saline for 30 min at 8 °C twice and the eluted proteins were precipitated using tricholoroacetic acid. In 293T cells, whole-cell extracts for denaturing immunoprecipitation were prepared by boiling the cells in lysis buffer (1% SDS, 20 mM Tris pH 8.0, 10 mM dithiothreitol), and the SDS concentration was adjusted to 0.1% by adding dilution buffer before adding the Flag antibody for immunoprecipitation. Nuclear extracts were prepared using lysis buffer (10 mM HEPES pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 1% NP40) to lyse cell membranes; the supernatant is the cytosolic fraction and the pellet is the nuclear fraction. For co-immunoprecipitation, the nuclear pellet was re-suspended in lysis buffer (0.1% NP40, 150 mM NaCl, 20 mM Tris pH 8.0 and 10% glycerol) and sonicated to disrupt the nuclei and chromatin. The antibody and nuclear extract were incubated overnight with 5 μg of CK2α antibody or 20 μl of M2 Flag antibody conjugated to magnetic beads. The beads were washed three times with the same lysis buffer, and the proteins bound to the beads were analysed by mass spectrometry or by immunoblotting.

Mass spectrometry. Protein samples were prepared as described[28]. In brief, the protein samples were diluted in TNE buffer (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA). RapiGest SF reagent (Waters Corp.) was added to the mix to a final concentration of 0.1% and samples were boiled for 5 min. TCEP (Tris (2-carboxyethyl) phosphine) was added to a final concentration of 1 mM and the samples were incubated at 37 °C for 30 min. Subsequently, the samples were carboxymethylated with 0.5 mg ml⁻¹ of iodoacetamide for 30 min at 37 °C followed by neutralization with 2 mM TCEP (final concentration). Proteins samples prepared as above were digested with trypsin (trypsin:protein ratio 1:50) overnight at 37 °C. RapiGest was degraded and removed by treating the samples with 250 mM HCl at 37 °C for 1 h followed by centrifugation at 23,000g for 30 min at 4 °C. The soluble fraction

was then added to a new tube and the peptides were extracted and desalted using C18 desalting columns (Thermo Scientific).

Trypsin-digested peptides were analysed by ultra-high-pressure liquid chromatography (UPLC) coupled with tandem mass spectroscopy (MS/MS) using nano-spray ionization as described[29]. The nano-spray ionization experiments were performed using a TripleTof 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase UPLC (Waters corporation nano ACQUITY) using a 20-cm, 75-micrometer inside diameter glass capillary tube packed with 2.5-μm C18 (130) CSH beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (acetonitrile) at a flow rate of 250 μl per min for 1 h. The buffers used to create the ACN gradient were: buffer A (98% H_2O, 2% ACN, 0.1% formic acid, and 0.005% TFA (trifluoroacetic acid)) and buffer B (100% ACN, 0.1% formic acid, and 0.005% TFA). MS/MS data were acquired in a data-dependent manner in which the MS1 (initial mass-to-charge-ratio (m/z) spectrum) data were acquired for 250 ms at an m/z of 400 to 1,250 Da and the MS2 (MS/MS or tandem MS) data were acquired from an m/z of 50 to 2,000 Da. The independent data acquisition parameters were as follows: MS1-TOF (time-of-flight) acquisition time of 250 ms, followed by 50 MS2 events

of 48 ms acquisition time for each event. The threshold to trigger an MS2 event was set to 150 counts when the ion had the charge state +2, +3 or +4. The ion exclusion time was set to 4 s. Finally, the collected data were analysed using Protein Pilot 4.5 (ABSCIEX) for peptide identifications.

Yeast strains and yeast plasmids used in the study. Yeast strains and yeast plasmids used in the study are described in Supplementary Table 4.

25. O'Neill, L. P. & Turner, B. M. Immunoprecipitation of native chromatin: NChIP. *Methods* **31,** 76–82 (2003).
26. Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11,** 83–93 (1997).
27. Kamps, M. P. & Sefton, B. M. Acid and base hydrolysis of phosphoproteins bound to immobilon facilitates analysis of phosphoamino acids in gel-fractionated proteins. *Anal. Biochem.* **176,** 22–27 (1989).
28. Guttman, M. *et al.* Interactions of the NPXY microdomains of the low density lipoprotein receptor-related protein 1. *Proteomics* **9,** 5016–5028 (2009).
29. McCormack, A. L. *et al.* Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69,** 767–776 (1997).
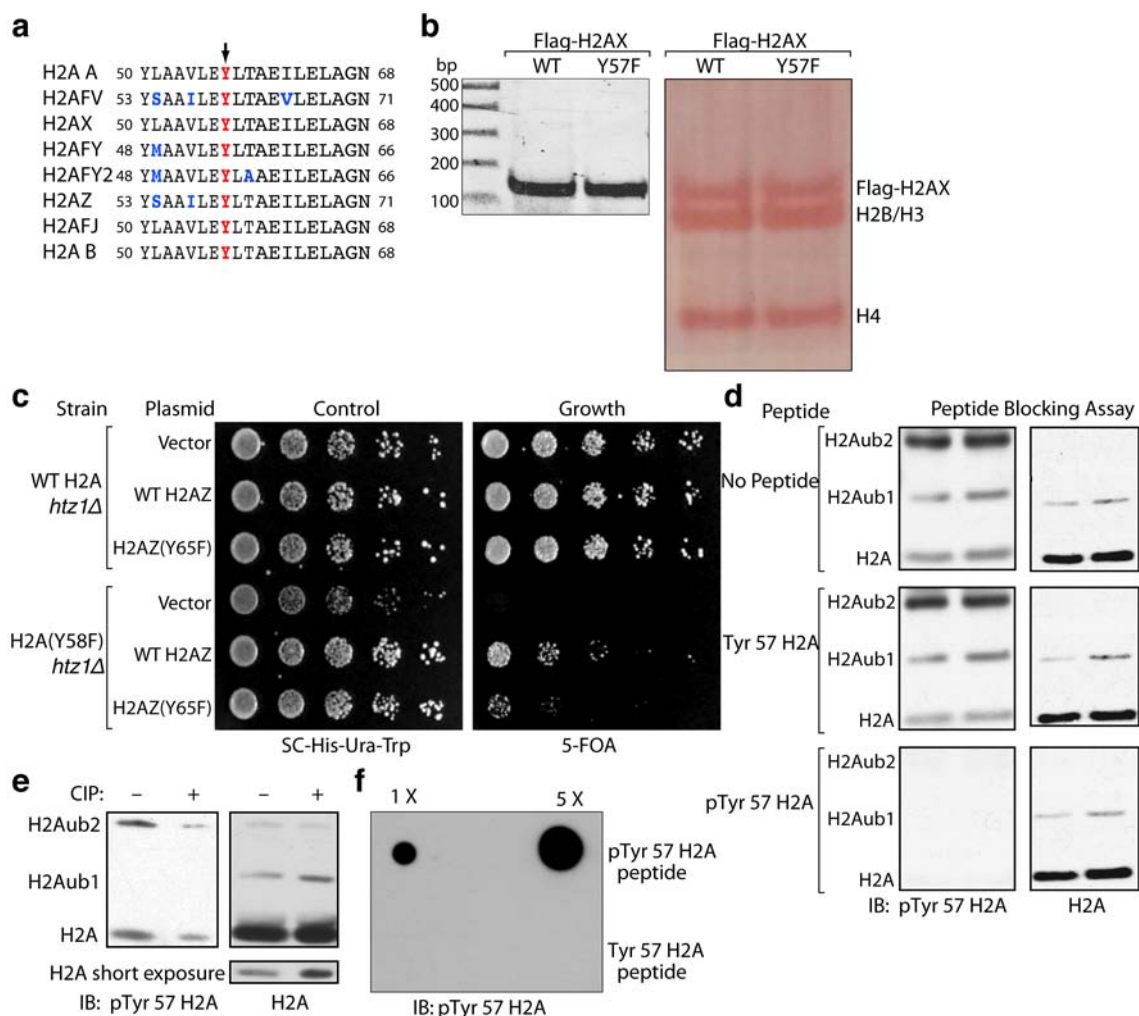
Extended Data Figure 1 | The conserved Tyr 57 residue in H2A is phosphorylated. a, The Tyr 57 residue is conserved in all variants of H2A. Sequences of H2A variants surrounding the Tyr 57 residue (arrow) in mammals is shown, particular variant residues are highlighted in blue. **b**, The Y57F mutation in H2A does not affect the structural integrity of nucleosomes. Mononucleosomes containing Flag-tagged wild-type H2AX or H2AX(Y57F) were immunoprecipitated, and histones and DNA were visualized by Ponceau staining (right) and by ultraviolet light (left), respectively. **c**, The H2A Tyr 58 residue has overlapping functions with the H2AZ Tyr 65 residue in yeast. Fivefold serial dilutions of the indicated transformants were plated on

SC–His–Ura–Trp for growth and 5-FOA for the loss of pJH33. **d**, **e**, H2A Tyr 57 is phosphorylated in 293T cells. **d**, Nuclear extracts from 293T cells were immunoblotted (IB) with anti-pTyr 57 H2A pre-incubated with indicated peptides, and re-probed with anti-H2A. **e**, Histone extracts from 293T cells were treated with calf intestinal phosphatase (CIP) for 1 h at 37 °C, and immunoblotted. **f**, The anti-pTyr 57 H2A antibody specifically recognizes the H2A peptide phosphorylated at Tyr 57 but not the non-phosphorylated peptide. Indicated peptides were spotted on nitrocellulose, and probed with anti-pTyr 57 H2A. Data represent three independent experiments.

a
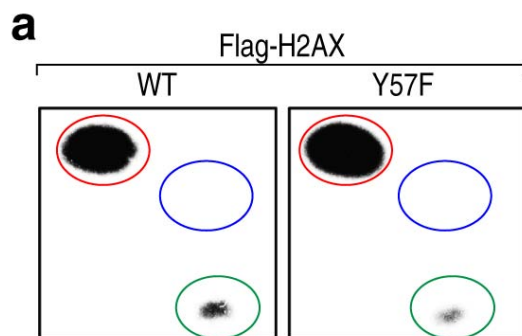
Extended Data Figure 2 | **CK2α phosphorylates Tyr 57 in H2A.** An *in vitro* kinase reaction was performed using recombinant GST-CK2α, 10 μCi of [γ-^{32}P]ATP supplemented with 10 μM cold ATP, and nucleosomes containing Flag-tagged wild-type H2AX or H2AX(Y57F) from 293T cells, and phosphoamino acid analysis of the phosphorylated Flag-tagged H2AX was performed. The red circle indicates pSer, the blue circle indicates pThr and the green circle indicates pTyr. Data represent two independent experiments.
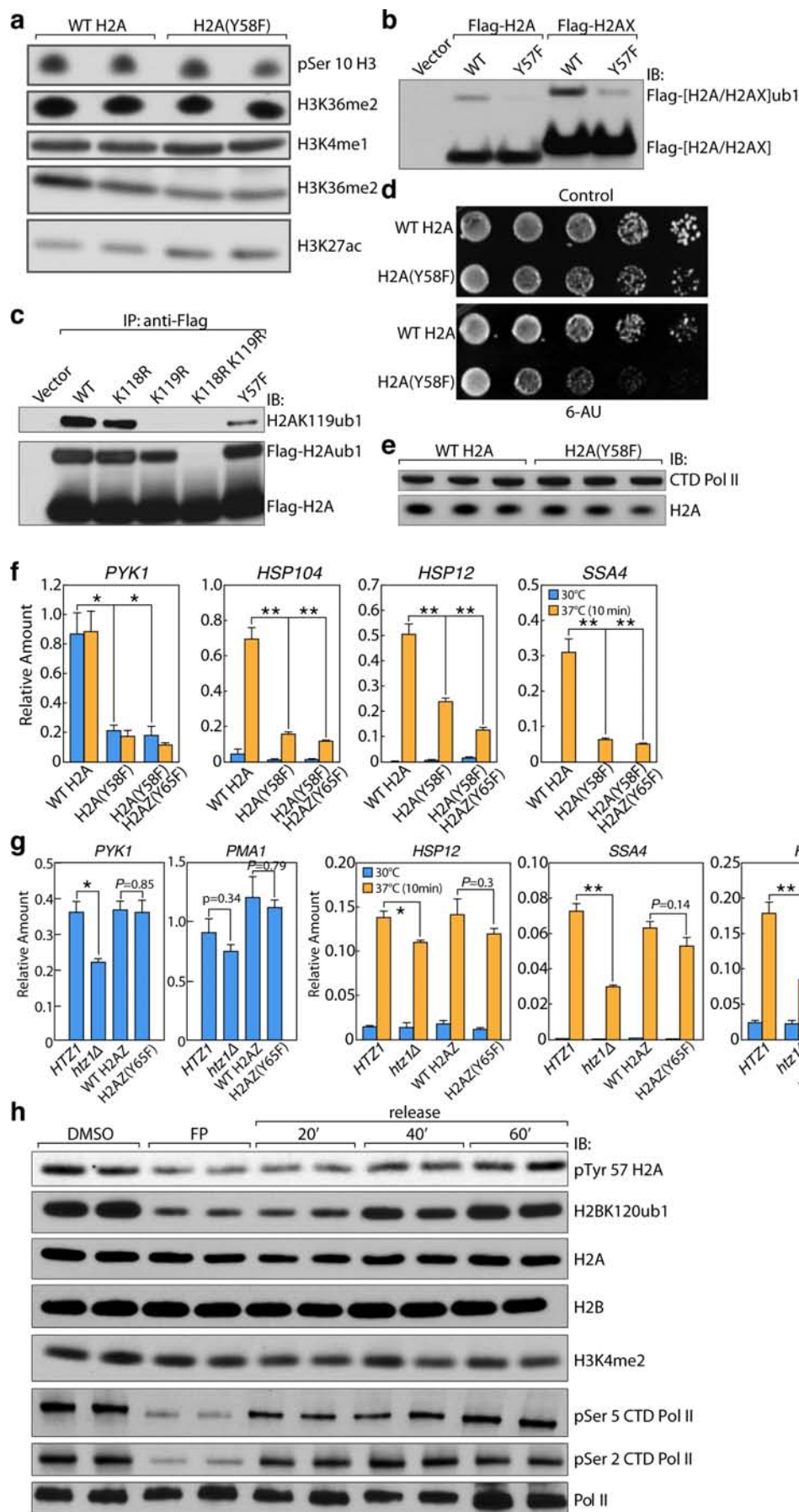
Extended Data Figure 3 | H2A Tyr 57 phosphorylation regulates transcriptional elongation. a, H2A(Y58F) mutation does not affect several other histone marks. Whole-cell extracts from wild-type (WT) or H2A(Y58F) yeast cells were immunoblotted (IB). **b, c**, H2A(Y57F) mutation affects H2A ubiquitination in 293T cells. **b**, Flag-tagged wild-type H2A/H2AX and Y57F mutants were expressed in 293T cells, and mono-ubiquitination was assessed by immunoblotting with Flag antibody. **c**, The Flag-tagged H2A mutants were expressed in 293T cells, immunoprecipitated (IP) under denaturing conditions, and immunoblotted. **d**, H2A(Y58F) mutant cells are defective in transcriptional elongation. Fivefold serial dilutions of wild-type and H2A(Y58F) cells were plated on SC supplemented with NH$_4$OH (solvent) or 100 μg ml^{-1} 6-azauracil (6-AU). **e**, Pol II protein level is comparable in wild-type and H2A(Y58F) yeast. Whole-cell extracts from wild-type or H2A(Y58F) yeast were immunoblotted. **f**, H2A(Y58F) mutation affects transcription. Wild-type H2A, H2A(Y58F), and H2A(Y58F) H2AZ(Y65F) yeast were grown at 30 °C or shifted to 37 °C for 10 min. RNA was extracted and transcript levels of the indicated genes were measured by reverse-transcription–qPCR, and normalized to *SCR1*, a Pol III transcript ($n = 3$, mean ± s.e.m., *$P < 0.05$, **$P < 0.01$). *P* values were calculated with two-tailed Student's *t*-tests. **g**, H2AZ(Y65F) mutation alone in yeast does not affect transcription significantly. Wild type (*HTZ1*) transformed with vector, and *htz1Δ* strains transformed with vector (*htz1Δ*), *HTZ1* (WT H2AZ) or *HTZ1(Y65F)* (H2AZ(Y65F)) were grown at 30 °C (blue bars) or shifted to 37 °C for 10 min. (orange bars), and transcript levels of the indicated genes were measured as in **f** ($n = 3$, mean ± s.e.m., *$P < 0.05$, **$P < 0.01$). *P* values were calculated with two-tailed Student's *t*-tests. **h**, Tyr 57 in H2A is phosphorylated during transcriptional elongation. 293T cells were treated with vehicle (DMSO) or flavopiridol (FP) (1 μM) for 4.5 h, then flavopiridol was washed out (release). Cells were harvested at the indicated minute (′) after release, and the nuclear extracts were immunoblotted. Data represent two (**a, d, h**) or three (**b, c, e–g**) independent experiments.
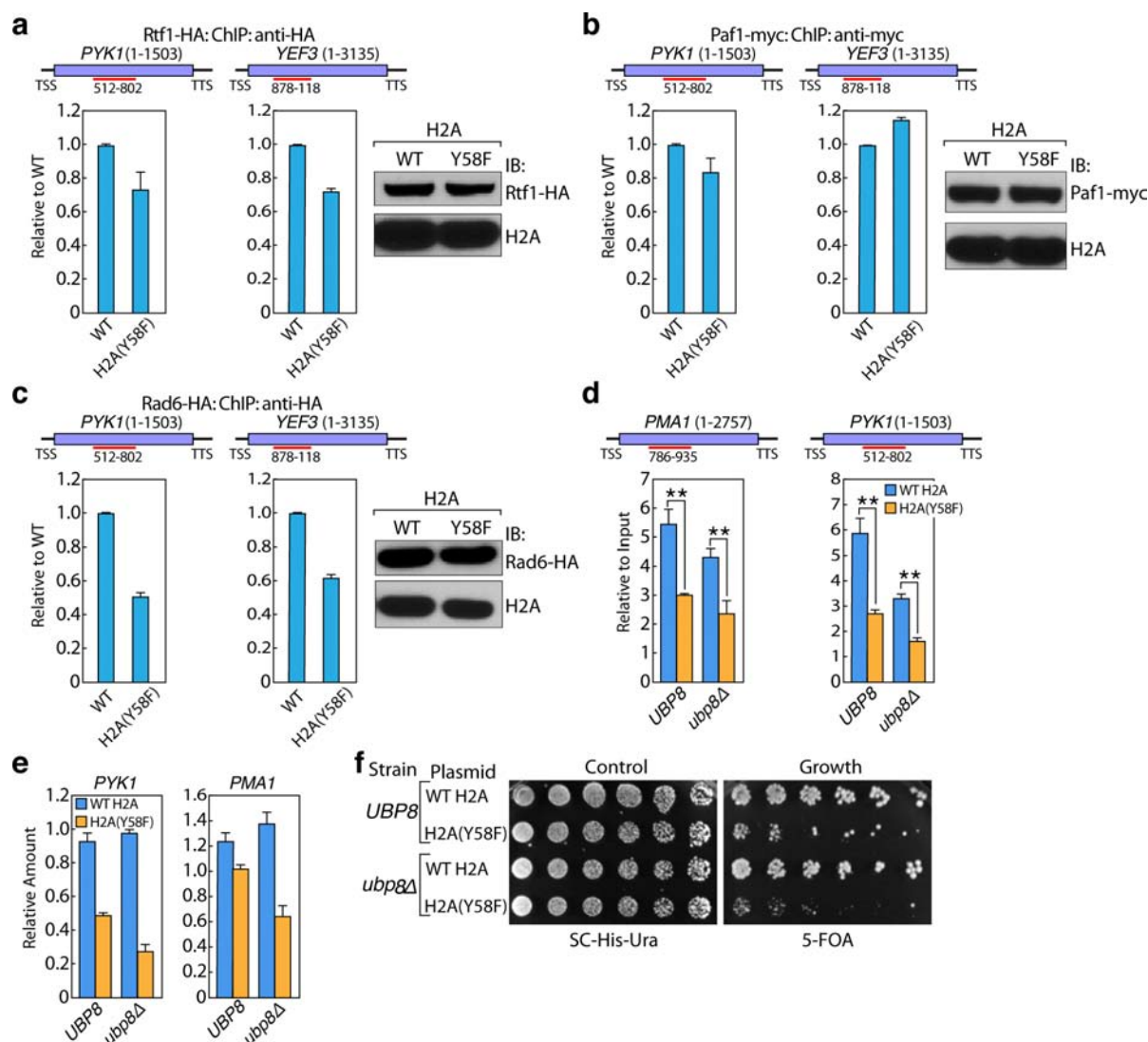
Extended Data Figure 4 | H2A(Y58F) mutation enhances H2B deubiquitination. a–c, The H2A Tyr 58 mutation has moderate to no effect on the recruitment of the H2B ubiquitination machinery. Binding of (**a**) Rtf1-HA, (**b**) Paf1-myc, and (**c**) Rad6-HA was measured by ChIP-qPCR in the indicated genes in wild-type and H2A(Y58F) yeast. Whole-cell extracts from the yeast strains were immunoblotted (IB) to compare the protein levels. ORF of the genes, and the region amplified by the primer pairs are shown ($n = 2$, mean ± s.e.m.). **d,** *UBP8* deletion does not rescue Pol II binding in the H2A(Y58F) mutant. Pol II binding in the indicated strains was measured by ChIP-qPCR. ($n = 3$, mean ± s.e.m., *$P < 0.05$, **$P < 0.01$). *P* values were

calculated with two-tailed Student's *t*-tests. The ORF of the genes and the regions amplified by the primer pairs are shown. **e,** *UBP8* deletion does not rescue the defect in transcriptional output in the H2A(Y58F) yeast. The mRNA levels of the indicated genes were determined by RT–qPCR and normalized to the *SCR1* transcript. ($n = 2$, mean ± s.e.m.). **f,** *UBP8* deletion does not rescue the growth defect in the H2A(Y58F) yeast. *UBP8* and *ubp8Δ* strains expressing either wild-type (WT) H2A or H2A(Y58F) were plated at 2.5-fold serial dilutions on SC–His–Ura for growth and 5-FOA for the removal of pJH33. Data represent two (**a–c, e, f**) or three (**d**) independent experiments.
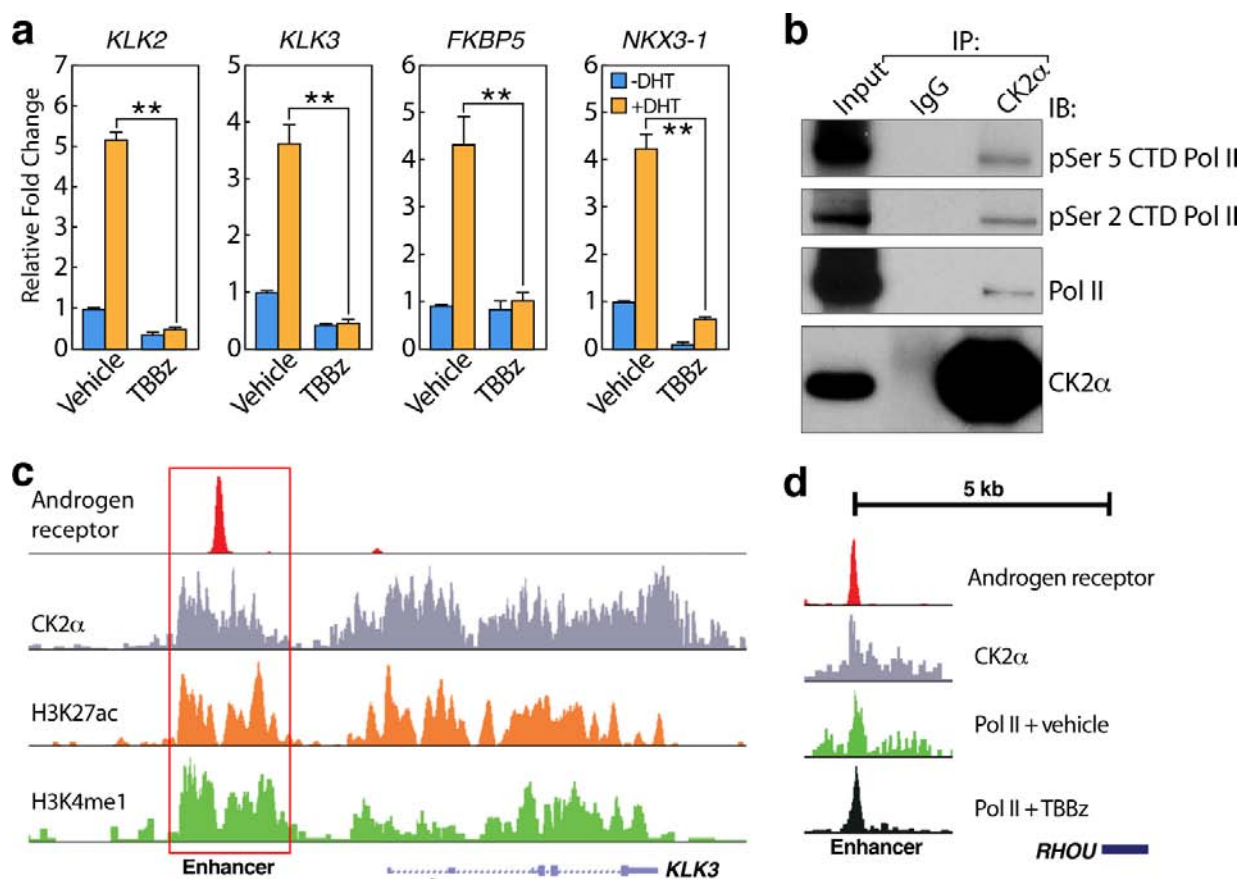
Extended Data Figure 5 | CK2 regulates transcriptional elongation. a, CK2 kinase activity is necessary for normal gene expression. LNCaP cells were treated with vehicle (DMSO) or TBBz (25 μM) for 60 min, and then treated with vehicle (ethanol) or 100 nM DHT for 90 min, and induction of the indicated androgen receptor (AR) target genes was measured by RT-qPCR ($n = 3$, mean ± s.e.m., $*P < 0.05$, $**P < 0.01$). P values were calculated with two-tailed Student's t-tests. **b**, Nuclear extracts from 293T cells were immunoprecipitated (IP) using CK2α antibody and immunoblotted (IB). **c**, Enrichment of CK2α, H3K4me1 (ref. 18), H3K27ac (ref. 18) and androgen receptor genes[18] at a representative androgen receptor enhancer (*KLK3*) is shown. **d**, Pol II tag density in cells treated with vehicle or TBBz at a representative *RHOU* enhancer is shown. Data represent two (**b**–**d**) or three (**a**) independent experiments; kb, kilobase.

LETTER

Regulation of RNA polymerase II activation by histone acetylation in single living cells

Timothy J. Stasevich[1,2,3], Yoko Hayashi-Takanaka[1,4,5], Yuko Sato[1,5], Kazumitsu Maehara[6], Yasuyuki Ohkawa[4,6], Kumiko Sakata-Sogawa[7,8], Makio Tokunaga[7,8], Takahiro Nagase[9], Naohito Nozaki[10], James G. McNally[11,12] & Hiroshi Kimura[1,4,5]

In eukaryotic cells, post-translational histone modifications have an important role in gene regulation. Starting with early work on histone acetylation[1], a variety of residue-specific modifications have now been linked to RNA polymerase II (RNAP2) activity[2,3], but it remains unclear if these markers are active regulators of transcription or just passive byproducts[4,5]. This is because studies have traditionally relied on fixed cell populations, meaning temporal resolution is limited to minutes at best, and correlated factors may not actually be present in the same cell at the same time. Complementary approaches are therefore needed to probe the dynamic interplay of histone modifications and RNAP2 with higher temporal resolution in single living cells[2,5,6]. Here we address this problem by developing a system to track residue-specific histone modifications and RNAP2 phosphorylation in living cells by fluorescence microscopy. This increases temporal resolution to the tens-of-seconds range. Our single-cell analysis reveals histone H3 lysine-27 acetylation at a gene locus can alter downstream transcription kinetics by as much as 50%, affecting two temporally separate events. First acetylation enhances the search kinetics of transcriptional activators, and later the acetylation accelerates the transition of RNAP2 from initiation to elongation. Signatures of the latter can be found genome-wide using chromatin immunoprecipitation followed by sequencing. We argue that this regulation leads to a robust and potentially tunable transcriptional response.

To monitor the impact of histone acetylation on transcription, we used a mouse cell line harbouring an inducible tandem gene array and expressing a green fluorescent protein (GFP)-tagged version of the glucocorticoid receptor (GFP–GR)[7]. Upon hormone stimulation, GFP–GR enters the nucleus and activates the array[7]. This leads to RNAP2 recruitment, initiation and elongation, as seen by immunofluorescence (Extended Data Fig. 1a, b) using antibodies against the RNAP2 carboxy-terminal domain (CTD) and its phosphorylation at serine 5 (Ser 5ph) and serine 2 (Ser 2ph)[8]. Immunofluorescence also suggested the array harbours high levels of histone H3 lysine-27 acetylation (H3K27ac) and H3K4 methylation (Extended Data Fig. 1b–d), markers of promoters and enhancers[2]. Of these markers, H3K27ac displayed the greatest variability between cells (Extended Data Fig. 1c), consistent with a rapid turnover[9,10] and a potential role in array regulation.

To visualize H3K27ac dynamics together with GR and RNAP2 phosphorylation in a single living cell, we prepared antigen-binding fragments (Fabs) conjugated to a fluorescent dye for use in Fab-based live endogenous modification labelling (FabLEM; Fig. 1a)[11,12]. Loaded Fabs accumulated in living nuclei and allowed us to track target modifications with a temporal resolution approaching 10 s (Extended Data Fig. 2). In accordance with this, Fabs against RNAP2 and H3K27ac quickly responded to gene activation at the array. As Fig. 1b shows, the array is hyper-acetylated at H3K27 before hormone stimulation, but levels

eventually drop after the arrival of GFP–GR and the start of RNAP2 elongation (Ser 2ph). The average result from 19 cells demonstrates acetylation levels remain high for about 5 min after elongation begins (Fig. 1c). To test the fidelity of our monitoring system, we treated cells with trichostatin A[13] or flavopiridol[14]. As expected, these drugs inhibited histone deacetylation as well as RNAP2 elongation at the array (Extended Data Fig. 3).

Taking advantage of the variability between cells of array H3K27ac levels before activation, we investigated how acetylation affects downstream transcription kinetics. We sorted cells from experiments into two groups of initially high and low array acetylation levels. If acetylation does not alter downstream transcription, we would expect no difference between measured GFP–GR or RNAP2 kinetics at the array from these groups. As Fig. 1c shows, however, there was a significant difference between groups. Namely, arrays with initially more histone acetylation had higher levels of downstream GFP–GR and elongating RNAP2 (Ser 2ph), a correlation (Extended Data Fig. 4a, b) that was independent of nuclear Fab concentrations (Extended Data Fig. 4c–e). In contrast, when we repeated the experiments with Fabs against initiated (Ser 5ph) rather than elongating (Ser 2ph) RNAP2 (Fig. 1d and Extended Data Fig. 3a), we found no difference in RNAP2 initiation levels between groups, although we confirmed the GFP–GR difference (Extended Data Fig. 4b). These data indicate that acetylation enhances both GR recruitment and RNAP2 elongation. This dual action seems to involve two temporally distinct mechanisms since the GR enhancement does not lead to enhanced RNAP2 initiation, presumably due to a slower step in-between. This prompted us to focus on the downstream enhancement of RNAP2 elongation, as this is the critical step in array transcription.

The H3K27ac-associated increase in RNAP2 elongation levels could be a negative effect due to slower RNAP2 elongation or it could be a positive effect due to more efficient RNAP2 promoter escape (hereafter we use the term 'promoter escape' as the transition to elongation indicated by the appearance of phosphorylated Ser 2). To distinguish these contradictory possibilities, we needed to quantify RNAP2 initiation separately from recruitment and elongation. For this, we performed a new set of FabLEM experiments to separately track RNAP2 recruitment (CTD), initiation (Ser 5ph) and elongation (Ser 2ph) at the array (Extended Data Fig. 5a, b). We renormalized data (summarized in Extended Data Fig. 5c–f) to indicate the number of RNAP2 at the gene array by independently measuring the number of RNAP2 per cell (Fig. 2a, b). We then fitted the data to a kinetic model of the RNAP2 transcription cycle[15] described in Fig. 2d. RNAP2 photobleaching data (Fig. 2c) was simultaneously fitted to enable accurate estimations of relatively fast model parameters (Extended Data Fig. 6a–c).

According to the simultaneous fit (reported below with 90% confidence intervals in parentheses; see Extended Data Table 1 for details),

**Figure 1 | Covalent modifications to histones and RNAP2 can be imaged in living cells using FabLEM. a**, Illustration of FabLEM strategy. **b**, Sample frames from a FabLEM experiment showing that the gene array (yellow arrow) is acetylated at H3K27 (H3K27ac) before activation by GFP–GR and accumulation of elongating RNAP2 (Ser 2ph). Scale bar, 5 μm. **c, d**, Fab recruitment curves showing the ratio of the average Fab intensity measured at the gene array to the average Fab intensity throughout the rest of the nucleus (Array/nuclear int. ± s.e.m.) from experiments tracking histone acetylation (blue, H3K27ac), GFP–GR (green) and RNAP2 elongation (red, Ser 2ph, $n = 19$, **c**) or RNAP2 initiation (purple, Ser 5ph, $n = 13$, **d**). H3K27ac levels drop (down arrows) about 5 min after RNAP2 initiation/elongation levels rise (up arrows). Data was split into two groups corresponding to cells with low (black) or high (grey) levels of H3K27ac at the array (relative to the nucleus) before gene activation.

the recruitment time Δ of RNAP2 to the gene array is ~2.3 min (Δ = 1.9–2.9 min). During this time, RNAP2 samples DNA relatively quickly, binding with rate k_{out} to potential sites on average for ~7 s ($1/k_{out}$ = 2–30 s). Of these sampling events, the rate of RNAP2 initiation k_{ini} is relatively low, so less than ~13% initiate ($k_{ini}/(k_{ini} + k_{out})$ = 2–30%). However, once initiated, the rate of promoter escape k_{esc} is much higher than the rate of abortion k_{ab}, so that ~90% produce transcript (with rate k_t) over a period of ~1.4 min ($k_{esc}/(k_{esc} + k_{ab})$ = 41–100%; $1/k_t$ = 1–2.5 min).

To explore how these kinetics depend on H3K27ac, we fitted data from sorted cells (Extended Data Fig. 6d–f). This revealed that RNAP2 promoter escape was ~0.75 min faster at arrays that initially had high levels of acetylation compared to low levels (Extended Data Table 1; $\Delta(1/k_{esc})$ = 0.65–0.86 min), a nearly 50% increase in speed (Extended Data Fig. 6f). All other parameters in the fit remained statistically similar (Extended Data Fig. 6e, f). For example, the next most significant change was in the fitted elongation time, which was ~0.27 min slower in cells with high levels of acetylation compared to low levels ($\Delta(1/k_t)$
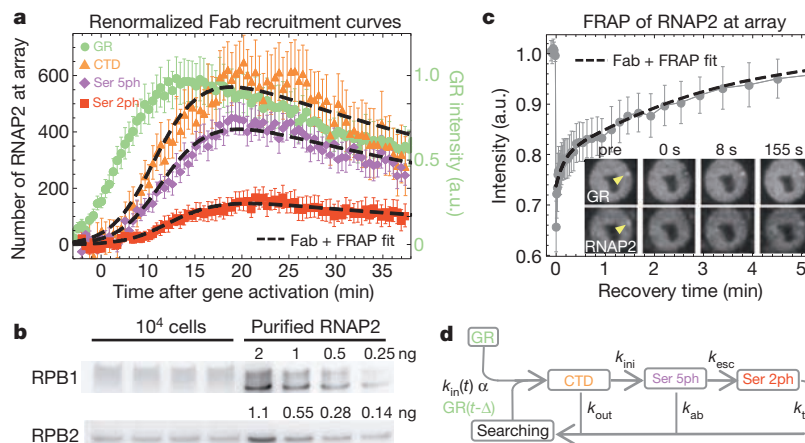
Figure 2 | Fitting the RNAP2 transcription cycle. a, RNAP2 Fab recruitment curves (± s.e.m., n = 12, 24 and 12 for CTD, Ser 5ph and Ser 2ph Fab, respectively) with the scale normalized to the number of RNAP2 at the gene array (note scale on right for GFP–GR data; $n = 24$; a.u., arbitrary units.). **b**, Example immunoblots used to estimate the number of RNAP2 per cell by comparing purified RNAP2 subunits (RPB1 and RPB2) with whole-cell extract from approximately 10^4 cells. **c**, The fluorescence recovery after photobleaching (FRAP) recovery curve (± s.e.m., $n = 19$) of mCherry-tagged RPB1 (RNAP2) at the gene array (inset, yellow arrow) with GFP–GR. When Fab and FRAP data are simultaneously fitted to the model in **d** (dashed lines in **a** and **c**), consistent parameter estimates are obtained (see Extended Data Table 1). **d**, RNAP2 transcription cycle model. Searching RNAP2 is recruited to promoters (CTD) with a binding off rate k_{out} and binding on rate $k_{in}(t)$ proportional to the amount of transcriptional activator (GR) at the gene a time Δ earlier. RNAP2 initiates (Ser 5ph) with rate k_{ini}, aborts with rate k_{ab}, escapes the promoter with rate k_{esc} (Ser 2ph) and terminates transcription with rate k_t.

= 0.19–0.34 min). The significance of the difference in promoter escape times was further validated by resampling of the data (Extended Data Fig. 6g, h). We therefore conclude that histone acetylation is positively correlated with transcription at the gene array, and this correlation seems to arise mainly by accelerating RNAP2 promoter escape rather than by altering elongation.

To test for a causal link between these events, we performed two sets of experiments to artificially lower array H3K27ac levels and quantify the impact of this perturbation on downstream transcription kinetics. First, we designed a construct containing the dimerization/DNA-binding domain of NF1A1.1 (nuclear factor 1; for array targeting[16]) followed by the H3K4 demethylase KDM5b[17,18]. When transiently transfected into cells, this construct demethylated array H3K4, leading to a loss of array H3K27ac (Fig. 3a, b), consistent with H3K4me3-mediated H3 acetylation[9]. After transcriptional activation, deacetylated arrays had less elongating RNAP2 (Ser 2ph) than control arrays, even though initiation levels (Ser 5ph) were unchanged (Fig. 3b). Data support a role for acetylation in promoter escape, although the methylation could also be contributing. To rule out the methylation, we next used the drug C646 (ref. 19) to selectively inhibit p300 and/or CBP, lysine acetyltransferases (KATs) known to mediate H3K27 acetylation[20]. In accordance with this, C646 treatment reduced array H3K27ac in 10 min without altering array H3K4me2 (Fig. 3c). Nevertheless, this again led to less elongating RNAP2 (Ser 2ph) than control arrays after activation, even though initiation levels (Ser 5ph) were actually slightly higher (Fig. 3d). Together these data indicate that array H3K27ac, rather than H3K4 methylation, promotes downstream RNAP2 promoter escape.

To probe the molecular origins of this effect, we next screened a panel of different histone-modifying enzymes to see if any colocalize to the array (Extended Data Fig. 7). Among the positive hits, we found p300 and CBP at the array (consistent with C646 data), along with intermittent accumulations of the deacetylases HDAC4 and HDAC7 (due to nucleocytoplasmic trafficking[21]; Extended Data Fig. 7a–c). The balance of KATs and HDACs at the array probably accounts for some of the variability of array H3K27ac levels. The confirmation of p300, in particular, prompted us to examine AFF4, a key component of the super

elongation complex (SEC)[22] that was recently shown to bind p300[23]. Consistent with this notion, we found that AFF4 is recruited less efficiently to lower-acetylated arrays in C646-treated cells than to control arrays after gene activation (Fig. 3e and Extended Data Fig. 7d). It is thus likely that AFF4 helps bridge H3K27ac to elongating RNAP2 via recruitment of the SEC.

To gauge the generality of our results, we performed a set of sequencing experiments. We first wanted to see if endogenous GR-response genes are also hyperacetylated before activation, as we observed at the array. To identify GR-response genes we sequenced RNA from cells before and after activation. Ranking genes by the fold increase in RNA identified many previously known GR-response genes (Supplementary Table). The top 1,000 genes had RNA levels increase between 2- and 33-fold, similar to the 6.4-fold increase at the gene array (Extended Data Table 1). To quantify H3K27ac enrichment at these genes we next sequenced DNA from chromatin-immunoprecipitation followed by sequencing experiments (ChIP-seq) using our H3K27ac antibody. This revealed the top 1,000 GR-response genes are indeed hyperacetylated at H3K27 before activation, as in the array (Fig. 4a).

We next examined whether the correlation between H3K27ac and RNAP2 promoter escape occurs more generally, so we performed an additional ChIP-seq experiment using our antibody against the RNAP2 CTD. This allowed us to categorize genes with similar levels of RNA expression by their H3K27ac content and to compare the distribution of RNAP2 at these genes (Extended Data Fig. 8). We found genes with higher levels of H3K27ac had more elongating RNAP2 downstream of the promoter relative to the amount at the promoter (Fig. 4b and Extended Data Fig. 8b). This implies that H3K27ac facilitates RNAP2 promoter escape at a broad set of endogenous genes, as our measurements at the gene array suggested.

In conclusion, we have shown in single living cells that H3K27ac regulates downstream transcription kinetics at two temporally distinct phases of the transcription cycle, indicating that the marker is more than a passive feature of gene expression networks[5]. H3K27ac can be thought of as a transcription gate-keeper, opening the entrance for incoming factors (GR recruitment) and opening the exit for outgoing factors (RNAP2
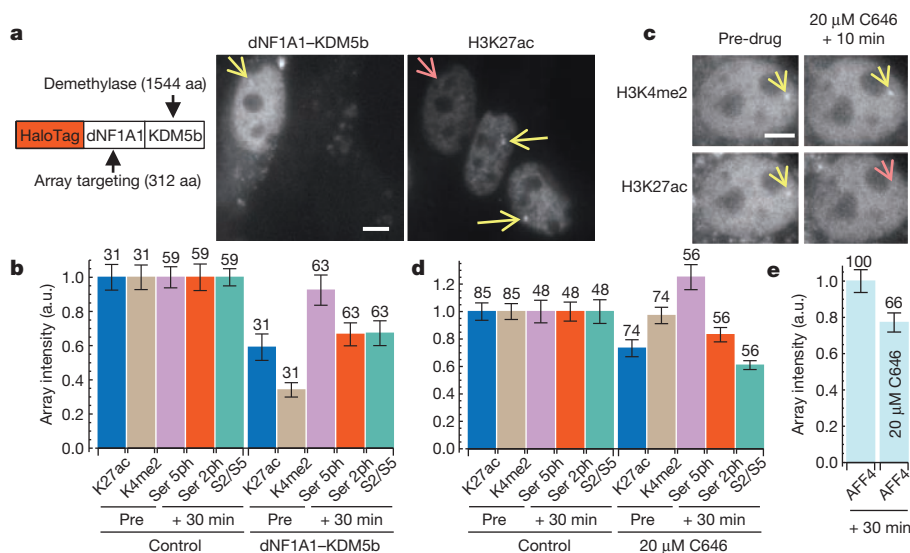
Figure 3 | Effect of perturbing array histone acetylation on RNAP2 transcription activation. a, A HaloTag construct that binds the array (via the DNA-binding domain of NF1A1.1, dNF1A1) and demethylates H3K4 (via KDM5b) to induce H3K27 deacetylation. When transfected into cells, HaloTag–dNF1A1–KDM5b accumulates at the array (yellow arrows) and H3K27ac levels decrease (pink arrow). aa, amino acids. Scale bar, 5 μm. **b**, Quantification of array histone and RNAP2 modification levels by immunostaining untreated cells (control) and cells transiently transfected with the dNF1A1–KDM5b construct. Cells were either fixed before (Pre) or 30 min

after (+ 30 min) transcriptional activation. S2/S5 is the ratio of array Ser 2ph to Ser 5ph; a.u., arbitrary units. **c**, An array (yellow arrow) in a living cell treated with 20 μM C646 for 10 min remains marked by Fab against H3K4me2, but is less marked by Fab against H3K27ac (pink arrow), indicating deacetylation. Scale bar, 5 μm. **d**, Same as **b**, but now in cells pretreated with vehicle (dimethyl sulphoxide; control) or with 20 μM C646 for 30 min. **e**, Quantification of array AFF4 levels in cells treated as in **d**. Error bars represent ± s.e.m. with the cell sample size reported above each bar.
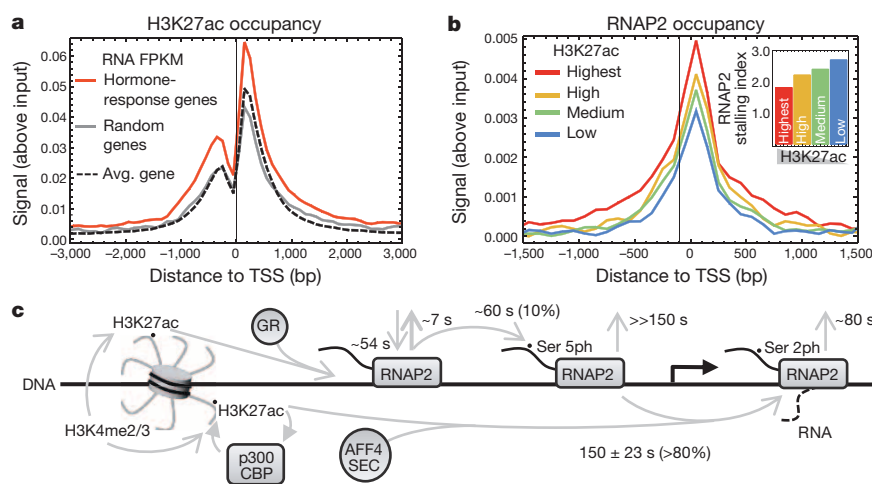
Figure 4 | Sequencing to examine genome-wide acetylation and transcription. a, Before activation, hormone-response genes ($n = 1,000$, red) have more H3K27ac compared to either random genes ($n = 1,000$, grey) expressing the same amount of RNA according to RNA-sequencing (2.5 fragments per kilobase of transcript per million mapped reads, FPKM) or the average gene (dashed). **b,** Genes with higher H3K27ac levels ($n = 2,000$) tend to have more RNAP2 and lower RNAP2 stalling indices (inset), that is, lower $\log_2$ ratios of promoter-bound RNAP2 (transcription start site TSS ± 350 base pairs (bp)) to elongating RNAP2 (from TSS +350 bp to +1,050 bp). **c,** A summary of measured RNAP2 transcription cycle kinetics and their regulation by H3K27ac.

promoter escape). This dual action enables a robust and potentially multi-faceted gene activation response. In our system the relatively slow recruitment of RNAP2 negated the upstream benefits of H3K27ac for GR accumulation, but this was rescued by late-acting acetylation that accelerated RNAP2 promoter escape. Presumably, in other cells with more efficient RNAP2 recruitment, the early- and late-acting acetylation could doubly enhance the transcriptional response. Thus, depending on acetylation levels and the cellular environment, the transcriptional response may be tunable. Natural variability of acetylation levels among cells seems to be driven by a balance of KATs and HDACs and mediated in part by H3K4 methylation. The resulting fluctuating levels can change RNAP2 promoter escape rates by as much as 50%. This sensitivity may explain why promoter escape at a hypoacetylated array[15,24,25] was far less efficient (<10%) than that which we observed at our hyperacetylated array. To guide future work, a model for the rapid response of GR-mediated gene activation that is consistent with all of our data is shown in Fig. 4c, along with a summary of fitted RNAP2 transcription kinetics. We believe the use of FabLEM combined with kinetic modelling will be a powerful tool for dissecting the dynamics of protein modifications in gene regulation more broadly. In particular, it will be interesting to see how other sets of genes that initially harbour more repressive marks are transcriptionally activated in single living cells.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12,** 599–606 (1998).
2. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nature Rev. Genet.* **12,** 7–18 (2011).
3. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).
4. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–412 (2007).
5. Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet.* **27,** 389–396 (2011).
6. Ecker, J. R. *et al.* Genomics: ENCODE explained. *Nature* **489,** 52–55 (2012).
7. McNally, J. G., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287,** 1262–1265 (2000).
8. Egloff, S., Dienstbier, M. & Murphy, S. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.* **28,** 333–341 (2012).
9. Crump, N. T. *et al.* Dynamic acetylation of all lysine-4 trimethylated histone H3 is evolutionarily conserved and mediated by p300/CBP. *Proc. Natl Acad. Sci. USA* **108,** 7814–7819 (2011).
10. Clayton, A. L., Hazzalin, C. A. & Mahadevan, L. C. Enhanced histone acetylation and transcription: a dynamic perspective. *Mol. Cell* **23,** 289–296 (2006).
11. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39,** 6475–6488 (2011).
12. Hayashi-Takanaka, Y., Yamagata, K., Nozaki, N. & Kimura, H. Visualizing histone modifications in living cells: spatiotemporal dynamics of H3 phosphorylation during interphase. *J. Cell Biol.* **187,** 781–790 (2009).
13. Smith, C. L. A shifting paradigm: histone deacetylases and transcriptional activation. *Bioessays* **30,** 15–24 (2008).
14. Chao, S.-H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription *in vivo. J. Biol. Chem.* **276,** 31793–31799 (2001).
15. Darzacq, X. *et al. In vivo* dynamics of RNA polymerase II transcription. *Nature Struct. Mol. Biol.* **14,** 796–806 (2007).
16. Stavreva, D. A. *et al.* Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nature Cell Biol.* **11,** 1093–1102 (2009).
17. Yamane, K. *et al.* PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol. Cell* **25,** 801–812 (2007).
18. Roesch, A. *et al.* A temporally distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141,** 583–594 (2010).
19. Bowers, E. M. *et al.* Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. *Chem. Biol.* **17,** 471–482 (2010).
20. Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J.* **30,** 249–262 (2011).
21. Yang, X.-J., Grégoire, S. & Class, I. I. Histone deacetylases: from sequence to function, regulation, and clinical implication. *Mol. Cell. Biol.* **25,** 2873–2884 (2005).
22. Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol. Cell* **37,** 429–437 (2010).
23. Byun, J. S. *et al.* ELL facilitates RNA polymerase II pause site entry and release. *Nature Commun.* **3,** 633 (2012).
24. Rafalska-Metcalf, I. U., Powers, S. L., Joo, L. M., LeRoy, G. & Janicki, S. M. Single cell analysis of transcriptional activation dynamics. *PLoS ONE* **5,** e10272 (2010).
25. Zhao, R., Nakamura, T., Fu, Y., Lazar, Z. & Spector, D. L. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nature Cell Biol.* **13,** 1295–1304 (2011).

METHODS

Cells and cell culture. Transcription activation experiments were performed on cells from mouse adenocarcinoma cell line 3617. These cells express a GFP-tagged version of the glucocorticoid receptor (GFP–GR) under the control of a tetracycline-repressible promoter. The cells also harbour a ~200 copy tandem gene array of a mouse mammary tumour virus/Harvey viral Ras (MMTV/vHa-Ras) reporter, as described previously[7]. Cells were grown in an incubator at 37 °C with 5% CO_2 in Dulbecco's modified Eagle medium (DMEM; Nacalai Tesque) supplemented with antibiotics (100 µg ml^{-1} streptomycin, 100 U ml^{-1} penicillin), 10% fetal bovine serum (FBS; Tissue Culture Biologicals), and tetracyclin (3 µg ml^{-1}). 12–18 h before experiments, the cell media was changed to phenol red-free DMEM (Nacalai Tesque) prepared as above, but with hormone stripped FBS (Life Technologies) and lacking tetracyclin. Transcription was activated at the gene array by addition of dexamethasone (100 nM), a ligand for GR that enables the molecule to translocate into the nucleus, where it can bind to the tandem gene array at 800–1,200 sites[7] within the long terminal repeat. HeLa cells used in immunoblotting experiments were grown as above, but in DMEM lacking tetracycline.

Antibody preparation. To generate monoclonal antibodies directed against the CTD of RBP1 (the catalytic subunit of the RNAP2 complex), as well as serine-5 and serine-2 phosphorylated forms, mice were immunized with synthetic peptides SYSPTSPSY SPTSPSYSPC, SYSPT(phospho-S)PSYSPTSPSYSPC and SYSPTSPSY(phospho-S) PTSPSYSPC, coupled to keyhole limpet hemocyanin. All handling of mice was approved by the Hokkaido University Animal Experiment Committee (approval number: 11-0109) and carried out according to guidelines for animal experimentation at Hokkaido University, where Mab Institute Inc. is located. Animals were housed in a designated pathogen–free facility at Hokkaido University. Mice were humanely euthanized via cervical dislocation by technically proficient individuals.

After generating hybridomas, clones were screened by ELISA using peptides listed in Extended Data Fig. 1a. Clones CMA601 (CTD), CMA603 (Ser 5ph), and CMA602 (Ser 2ph) all reacted specifically with target levels of RPB1. Using a kit (AbD Serotec), clone CMA601 was isotyped as IgG1κ, while clones CMA603 and CMA602 were isotyped as IgG1λ. For antibody purification, hybridomas were grown in CD Hybridoma medium (Invitrogen) supplemented with 2 mM glutamine. The culture supernatant (250 ml) was then filtrated through a 0.20 µm pore filter and NaCl was added to a final concentration of 4 M. The supernatant was then filtered through a HiTrap Protein A FF Sepharose column (1 ml; GE Healthcare). After washing the column with Protein A IgG1 binding buffer (Thermo Fisher Scientific), IgG was eluted using Mouse IgG1 Mild Elution Buffer (Thermo Fisher Scientific) and concentrated up to 4–8 mg ml^{-1} in PBS using an Amicon Ultra filter (50K cut-off; Millipore). Monoclonal antibodies specific to histone modifications were prepared similarly, as previously described[11,12,26]. Antibodies against RPB2 (sc-55039; Santa Cruz Biotechnology), p300 (NM11; ab3164; Abcam), CBP (D6C5; Cell Signaling), SRC1 (128E7; Cell Signaling), and AFF4 (HPA023690; Atlas Antibodies) were purchased separately.

Antigen-binding fragment (Fab) preparation and fluorescence conjugation. From monoclonal antibodies, Fab were prepared using a kit (Thermo Fisher Scientific; Pierce Mouse IgG1 Fab and F(ab′)2 Preparation Kit) according to the manufacturer's instructions. The buffer was replaced with PBS and concentrated up to ~1 mg ml^{-1} using an Ultrafree 0.5 filter (10K cut-off; Millipore). The purity and integrity of Fab were analysed by SDS–PAGE using a 10–20% gradient gel (Wako). Purified Fab or IgG were conjugated with a fluorescent dye using Alexa 488 tetrafluorophenyl ester (Invitrogen), Cy3 N-hydroxysuccinimide ester (GE Healthcare), or Cy5 N-hydroxysuccinimide ester (GE Healthcare). Dried fluorescent dye esters (for labelling 1 mg protein) were dissolved into 50 µl (Alexa488) or 100 µl (Cy3 and Cy5) dimethyl sulphoxide (DMSO; Wako) and stored at −20 °C. Fab (100 µg) was diluted into 100 mM NaHCO$_3$ (pH 8.3) in 100 µl. After addition of a dye solution (5, 1.3 and 4 µl for Alexa488, Cy3 and Cy5, respectively), the mixture was incubated for 1 h at room temperature with gentle rotation. The sample was passed through a PD-mini G-25 desalting column (GE Healthcare), pre-equilibrated with PBS, to remove unconjugated dyes, and dye-conjugated Fab was concentrated up to ~1 mg ml^{-1} using an Ultrafree 0.5 filter (10K cut-off; Millipore). The Fab concentration and dye:protein ratio were calculated from the absorbance at 280 and 494, 552 or 650 nm, using the extinction coefficient of IgG and correction factor at 280 nm provided by the manufacturers (that is, 0.11, 0.08 or 0.05). Fluorescent dye-labelled Fab samples that yielded dye:protein ratios 0.5–2 were used for live imaging.

Loading fluorescent Fabs into living cells. Cells were plated on a glassbottom dish with a coverslip (Mat-Tek) and the next day fluorescent Fab was loaded into cells using a bead-loading method[11,12,27,28], as follows. The medium was removed from the dish and saved, fluorescent Fab was pipetted onto the coverslip centre (1 mg ml^{-1} in PBS; 2–4 µl), and glass beads (106 µm; Sigma-Aldrich; G-4649) were sprinkled on top. After tapping the dish four to eight times, the original DMEM was added back to the dish and the cells were returned to an incubator for 1–2 h. Cells were then trypsinized and plated into an 8-well Labtek II microscope chamber (Nalgene) with phenol red-free DMEM (Nacalai Tesque) containing hormone-stripped FBS

(Life Technologies) and lacking tetracycline. After an additional 12–18 h, time-lapse imaging or FRAP experiments were performed. Loaded cells remained healthy, continued to divide at the same rate as unloaded cells, and had the same transcriptional activity (indicated by Ser 2ph) as unloaded cells (Extended Data Fig. 4c, d).

Time-lapse imaging of Fab recruitment to the tandem gene array. Using a confocal microscope (FV-1000; Olympus) equipped with a PlanSApo ×60 (NA = 1.40) oil-immersion objective and a cell culture system (Tokai Hit) set at 37 °C with 5% CO_2, multicolour image stacks (0.3–0.6% of 20 mW 488 nm laser, 17–25% of 2 mW 543 nm laser, 0.5–0.9% of 5 mW 633 nm laser) were acquired in sequential line imaging mode (3-line Kalman) using filters for EGFP (BA505–525 nm), Cy3 (BA560–620) and Cy5 (BA650IF). After addition of dexamethasone to the cell media (100 nM), 100 image stacks were acquired (256 × 256 × 6 voxels; 0.092 µm × 0.092 µm × 0.8 µm per voxel; 2 µs per pixel; 0.42 min per stack, pinhole open). The z-range was chosen to span the majority of the cell nucleus and include the entire tandem gene array. Image stacks were aligned and the average fluorescence intensity of the nucleus as well as the total and the average fluorescence intensity of the tandem gene array were measured from all slices of each stack using custom Mathematica code (Supplementary Information; Wolfram Research), (Extended Data Fig. 5c). The recruitment curves for each channel were calculated as the ratio of the average array intensity to the average nuclear intensity to correct for photobleaching and/or changes that might occur in the nucleus at a global level (for example, changes associated with the cell cycle). Noisy data or data from stacks that could not be aligned were discarded. Average recruitment curves from all cells were aligned to the activation time of the array by lining up phenomenological fits to the GFP–GR recruitment curves, as detailed in the section below on data fitting and shown in Extended Data Fig. 9a. Fab experiments were repeated on cells prepared independently on at least three separate days. Reported error bars represent ± s.e.m.

Plasmid construction and transient transfection. To construct an expression vector of mCherry–RPB1, complementary DNA was excised from eGFP–RPB1 vector[29] using NheI, blunted, and ligated into pmCherry-C1 that was EcoRI-digested and blunted. The resulting plasmid was verified by nucleotide sequencing. To construct an expression vector of dNF1A1–KDM5B, the first 308 N-terminal amino acids of NF1A1.1 (containing the dimerization and DNA-binding domains[30]) were amplified from mCherry–NF1A1.1[16] by PCR (PrimeStar; TaKaRa) with flanking SgfI sites (GCGATCGC) using the following primers: 5′-ATGCGATCGCCGAT GAGTTTCATCCTTTCATTGAAG-3′ (Forward) and 5′-ATGCGATCGCACCA GGACTGTCCATTTC-3′ (Reverse). The resulting products were purified (QIAquick Gel Extraction Kit; Qiagen) and inserted into the SgfI site of HaloTag–KDM5b (Kazusa DNA Research Institute; Flexi HaloTag clone FHC27753). mCherry–RPB1, mCherry–H2B[31,32] and all HaloTag-tagged proteins (Extended Data Fig. 7; Kazusa DNA Research Institute; Flexi HaloTag clones) were transiently transfected into cells with Opti-mem medium (Invitrogen) using the Lipofectamine 2000 reagent (Invitrogen), according to the manufacturer's instructions. After an incubation time of ~4 h, the media was changed to a phenol-red-free medium to eliminate background fluorescence and cells were examined >12 h later.

Fluorescence recovery after photobleaching. FRAP was performed on a confocal microscope (FV-1000; Olympus) equipped with a PlanSApo ×60 (NA = 1.40) oil-immersion objective and a cell culture system (Tokai Hit) set at 37 °C with 5% CO_2. For mCherry–RPB1/GFP–GR FRAP (Fig. 2c), 20 multicolour images were collected in a sequential line-scan fashion (0.3% of 20 mW 488 nm laser, 25% of 2 mW 543 nm laser; 128 × 64 pixels; 0.232 µm per pixel; 0.703 s per frame; 20 µs per pixel; pinhole open) using filters for eGFP (BA505–525 nm), and Cy3 (BA560IF). A circular area 7 pixels in diameter was then photobleached (100% 488 nm, 543 nm laser transmission, single iteration 50 ms), and a further 80 frames were collected. Following this, an additional 100 frames were acquired with the same settings but at a rate of 12 frames per min. Control FRAP experiments were performed to demonstrate no reversible photobleaching[31] of mCherry–RPB1 (Extended Data Fig. 9b) as well as no role of diffusion[33] in the mCherry–RPB1 recovery beyond 0.7 s post-bleach (Extended Data Fig. 9c).

For FRAP experiments on Alexa-488-conjugated Fab against RNAP2 (Extended Data Fig. 2), 9 images were collected (1% 488 nm laser transmission; 128 × 64 pixels; 0.209 µm per pixel; 106 ms per frame; 2 µs per pixel; pinhole open), a circular area 9 pixels in diameter was photobleached (100% 488 nm laser transmission, single iteration 100 ms), and a further 791 frames were collected. All frames from FRAP movies were aligned (when required) and the average intensity in the bleached area was measured using custom Mathematica code (Supplementary Information; Wolfram Research). Noisy data or data from stacks that could not be aligned were discarded. Unintentional photobleaching was corrected for by normalizing the total nuclear intensity to its original value. The corrected curves were then fit to a reaction-diffusion equation to extract Fab diffusion coefficients (D), bound fractions (BF), and binding times (t_{off}), as described previously[34]. The diffusion coefficient of Fab in the absence of binding was determined to be 20 ± 8 µm^2 s^{-1} from line-scan FRAP experiments on Fab prepared from an anti-DYKDDDDK (Flag) monoclonal

antibody (Wako) bead-loaded into the same cells. This Fab displays purely diffusive recovery as there are no target FLAG sequences in the cells to which it could bind (Extended Data Fig. 2b). In line-scan FRAP[34], a single line (1.4% 488 nm; 128×1 pixels; 2 µs per pixel; pinhole open) passing through the centre of a 9 pixel diameter photobleach spot (75% 405 nm laser transmission; 23 ms) was continuously scanned to quantify the bleach spot profile with time. In total there were 21,000 scans, with the bleach occurring between scans 100–116. Unintentional photobleaching was corrected by repeating the experiment without the photobleach, either before the original photobleach experiment or after (the order of these two experiments had no effect on the results). FRAP experiments were repeated on cells prepared independently on at least three separate days. Reported error bars represent ± s.e.m.

Immunofluorescence. Cells grown on coverslips were fixed with 4% formaldehyde (Electron Microscopy Sciences) in 250 mM HEPES–NaOH (pH 7.4) for 10 min at room temperature, washed three times with PBS, and incubated for 2 h at room temperature in 40 µl PBS containing 10% Blocking One-P (Nacalai Tesque), 0.5% Triton X-100, and 2 µg ml^{-1} Cy3/Cy5-conjugated IgG. After washing three times with PBS, coverslips were mounted using Prolong-Gold (Invitrogen). Fluorescence images were sequentially collected using a confocal microscope (FV-1000; Olympus) with the same settings as for time-lapse imaging described above, but with a pixel size of 0.046 µm. For immunostaining data in Fig. 3 and Extended Data Figs 1c, 4c and 7b, d, images were collected using an electron multiplying charge-coupled device (iXon+; Andor; normal mode; gain 5.1; exposure period 100–1,000 ms) installed on a Nikon Ti-E widefield microscope equipped with a ×100 PlanApo VC objective (N.A. 1.4), a 75 W Xenon lamp illumination, and filter sets (Semrock; LF488-A for GFP, LF561-A for Cy3/TRITC and Cy5-4040A for Cy5). For quantification purposes, experiments were repeated on cells prepared independently on at least two separate days. Reported error bars represent ± s.e.m.

Determining the number of RNAP2 molecules per cell. The procedure involved three steps and is summarized in Fig. 2b. In step one, a fixed number of cells were obtained as follows. Cells in four 10-cm dishes were grown under identical conditions to near confluency. In two dishes, cells were counted to ensure the consistency of growth conditions and the average number of cells per dish was calculated. The remaining two dishes were washed three times with serum-free media, all media was removed, and cells were scraped off each dish for transfer to Eppendorf tubes. These were boiled for 10 min at 95 °C and 2 × SDS gel loading buffer was added to each Eppendorf tube to adjust the final concentration to ~5×10^6 cells ml^{-1}. This process was repeated once more and samples were saved for whole-cell extract gels in step three. In step two, RNAP2 protein was immunoprecipitated from cells grown in 16–24 15-cm dishes using our monoclonal antibody against the RNAP2 CTD. A 5 µl aliquot of the immunoprecipitate was resolved on gels with Coomassie brilliant blue in comparison to approximately 4, 12, 37, 111, 333 and 1,000 ng of albumin protein (96% pure, Nacalai Tesque). After digital images were collected, a standard curve relating loading to signal intensity was constructed from the albumin intensities. By linear interpolation to the standard curve, the amount of RNAP2 protein in the aliquot was calculated (see Extended Data Fig. 9f for an example of how this interpolation was done). In the third step, aliquots of the immunoprecipitate containing 2, 1, 0.5, 0.25 and 0.125 ng of RNAP2 along with four whole-cell extract aliquots containing 10^4 cells from the first step were resolved on gels. These were then immunoblotted and, as in step two, digital images were collected and a standard curve was constructed from the RNAP2 intensities. By linear interpolation to the standard curve, the number of protein per cell was calculated. This was repeated at least two times, yielding a value of 200,000 ± 30,000 (± s.e.m.) RNAP2 per cell of the 3617 cell line. This value was confirmed by repeating the analysis, but now with an antibody against a different subunit of RNAP2, RPB2 (which yielded a value of 220,000 ± 30,000).

The measurement above was further corroborated by an independent measurement, summarized in Extended Data Fig. 9e, f. Whole-cell extract from 3617 cells and HeLa cells were prepared as in step one above, and dilution series containing approximately 5, 1.6, 0.5, 0.185 and 0.0617 × 10^4 cells were resolved on gels. These were then blotted on filter paper, and the filter paper probed with specific RNAP2 monoclonal antibodies (CTD, Ser 5ph and Ser 2ph). Standard curves were created from one of the dilution series and the relative amount of the remaining series was calculated by linear interpolation to the standard curve. Based on this and our previous measurement of 320,000 ± 58,000 (± s.e.m.) RNAP2 molecules per HeLa cell[35], the number of RNAP2 molecules per 3617 cell was calculated as 170,000 ± 40,000 RNAP2 per 3617 cell. The average of this estimate and the estimate in the previous paragraph was used as our final estimate of 185,000 ± 20,000 RNAP2 per cell reported in Extended Data Table 1.

Determining the number of phosphorylated Ser 2/Ser 5 of RNAP2 per cell. To calculate the number of RNAP2 phosphorylated at Ser 5 and Ser 2, immunoblots (as shown in Extended Data Fig. 9e) were further analysed by comparing the relative size of the α-CTD, α-Ser 5ph, and α-Ser 2ph blots. α-CTD lanes displayed the widest band, consistent with the ability of the α-CTD antibody to bind to all forms of RNAP2. This wide band was bounded by relatively dark upper and lower borders

that correspond to maximally phosphorylated and unphosphorylated RNAP2, respectively. In contrast, α-Ser 2ph lanes displayed only a single upper band that corresponds to highly phosphorylated RNAP2, while α-Ser 5ph lanes displayed a somewhat wider band that encompassed the α-Ser 2ph band, but did not extend down as low as the α-CTD band (that is, contained all forms of RNAP2 except for the fully unphosphorylated form). Examples are shown in Extended Data Figs 5f and 9e.

The overlapping α-Ser 5ph and α-Ser 2ph bands is consistent with the observation that highly phosphorylated RNAP2 contains both Ser 2 and Ser 5 phosphorylation[36–38]. As the α-Ser 5ph antibody can bind to Ser 5ph (Extended Data Fig. 1a), RNAP2 that is phosphorylated at both Ser 5 and Ser 2 can be detected. This is corroborated by three other independent experiments. First, immunoprecipitated Ser 5ph RNAP2 could be stained with the α-Ser 2ph antibody (data not shown), demonstrating directly that RNAP2 can have some repeats in its C-terminal domain that are phosphorylated at Ser 5 and others that are phosphorylated at Ser 2. Second, ChIP-chip (chromatin immunoprecipitation on chip) data[36] has shown that Ser 5ph is still present downstream of promoters, suggesting the mark remains for a significant amount of time while Ser 2ph proceeds. And third, novel binding motifs have recently been identified in histone methyl-transferase Set2 that preferentially bind to heptad repeats in the C-terminal domain of RNAP2 phosphorylated at both Ser 2 and Ser 5[37,38]. Given this evidence, we calculated the fraction of total RNAP2 phosphorylated at Ser 2 (or Ser 5) as the fraction of overlap between the α-Ser 2ph (or α-Ser 5ph) band and the α-CTD band (Extended Data Fig. 5f). This yielded 62,000 ± 11,000 (± s.e.m.) RNAP2 phosphorylated at Ser 2 per 3617 cell and 123,000 ± 15,000 RNAP2 phosphorylated at Ser 5 per 3617 cell (the latter reported in Extended Data Table 1).

We independently verified the Ser 2ph fraction by doing FRAP experiments on cells transiently transfected with mCherry–RPB1 and treated or not treated with the Ser 2 phosphorylation inhibiting drug flavopiridol (1 µM, 1 h), as shown in Extended Data Fig. 9d. Treated cells recovered 21% more than untreated cells (with the recovery baseline determined by analogous FRAP experiments in cells transiently expressing mCherry–H2B), indicating at least 21% of RNAP2 has Ser 2ph (39,000 ± 7,000). We used the average of the two independent estimates as the number of RNAP2 phosphorylated at Ser 2 per 3617 cell: 49,000 ± 7,000, reported in Extended Data Table 1 and shown in Extended Data Fig. 5f.

Renormalizing FabLEM data to number of RNAP2 at the array. The workflow is summarized in Extended Data Fig. 5c–f. In detail, to renormalize RNAP2 FabLEM data (expressed as the mean array to nuclear signal S, as shown in Fig. 1c, d and Extended Data Fig. 5a, b) to absolute RNAP2 number, the total RNAP2 Fab nuclear signal was equated to the total number of RNAP2 per cell (calculated by quantitative immunoblotting, as described above). However, the fraction FF of freely diffusing, unbound Fab was first subtracted because this part of the signal does not represent target RNAP2. FF was determined for each Fab from quantitative FRAP experiments (Extended Data Figs 2 and 5e) measuring the total Fab bound fraction, BF$_{tot}$, from which FF = 1 − BF$_{tot}$. Also, as S is the mean array-to-nuclear intensity per pixel, total array to total nuclear intensity were calculated by multiplying S (after subtracting FF) by the average array/nuclear volume V_{arr}/V_{nuc} (measured by confocal 3D image stacks as shown in Extended Data Fig. 5d and reported in Extended Data Table 1).

To demonstrate this, let I_{arr} be the mean intensity of the array Fab signal, I_{nuc} be the mean intensity of the nuclear Fab signal, and BG be the portion of the signal attributed to freely diffusing Fab (that is, the background). Then the number of RNAP2 at the array n_{arr} is proportional to the background subtracted mean intensity multiplied by the average size of the array:

$$n_{arr} \propto (I_{arr} - BG) V_{arr} \tag{1}$$

where V_{arr} is the volume of the array. Likewise for the number of RNAP2 in the nucleus n_{nuc}:

$$n_{nuc} \propto (I_{nuc} - BG) V_{nuc} \tag{2}$$

where V_{nuc} is the volume of the array. Thus,

$$\frac{n_{arr}}{n_{nuc}} = \frac{I_{arr} - BG}{I_{nuc} - BG} \frac{V_{arr}}{V_{nuc}} \tag{3}$$

We estimate BG as the mean Fab nuclear signal I_{nuc} multiplied by the fraction of Fab freely diffusing FF: BG = FF $\times I_{nuc}$. Thus, we have

$$\frac{n_{arr}}{n_{nuc}} = \frac{I_{arr} - I_{nuc}FF}{I_{nuc} - I_{nuc}FF} \frac{V_{arr}}{V_{nuc}} \tag{4}$$

Simplifying and solving for n_{arr} gives the following formula for renormalization:

$$n_{arr} = n_{nuc} \frac{I_{arr}/I_{nuc} - FF}{1 - FF} \frac{V_{arr}}{V_{nuc}} \tag{5}$$

Noting that $S \equiv I_{arr}/I_{nuc}$ yields the renormalization formula shown in Extended Data Table 1.

Immunoprecipitation. Cells from a 10-cm dish were spun down and washed three times in a physiological buffer[35] (PB; 100 mM potassium acetate, 20 mM KCl, 10 mM Na$_2$HPO$_4$, 1 mM MgCl$_2$, 1 mM disodium ATP, 1 mM dithiothreitol), resuspended in ~1 ml PB$^+$ (PB with 0.3 M NaCl, 0.1% Triton, and Roche protease and phosphatase inhibitor cocktails), transferred to an Eppendorf tube, and gently rotated for 30 min at 4 °C. Cells were then spun down at 4 °C for 5 min at 5,000g and the supernatant mixed with 250 µl Dynabeads (Life Technologies, M-280; sheep anti-mouse IgG) that had previously been soaked for 4 h in 500 µl of PBS containing 10 µg of mouse IgG against RNAP2 and subsequently equilibrated again with PB$^+$. The pellet was resuspended in 1 ml PB$^+$, sonicated at 4 °C for 30 min (repeating a 30 s on, 30 s off cycle), spun down at 4 °C for 20 min at 20,000g, and the supernatant collected and mixed with Dynabeads, prepared as above. The Dynabead mixtures were then rotated overnight, washed four times with PB$^+$, one time with PB, and spun down on a table-top centrifuge so that all media could be removed aside from the beads. Proteins were eluted from beads by adding 20 µl 2 × SDS gel loading buffer followed by denaturing at 95 °C for 10 min. Samples were spun down to separate beads and the supernatant with eluted protein sucked out and placed in a new Eppendorf. This was then placed on a magnetic rack to further separate the remaining beads and the supernatant containing eluted protein was again sucked out and placed in a new Eppendorf from which aliquots for immunoblotting could be taken, as described below.

Immunoblotting. Samples of denatured proteins mixed with 2 × SDS gel loading buffer were resolved on SDS-polyacrylamide gels, and blotted onto polyvinylidene difluoride filters (Pall). The filters were washed in TBST (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.05% Tween 20), blocked for 30 min with Blocking One-P (Nacalai Tesque), incubated for 2 h with primary antibody (0.2–1 µg ml^{-1}) diluted in Immuno Enhancer Solution 1 (Toyobo), washed three times in TBST over 15 min, incubated for 2 h with a 1 in 5,000 to 1 in 2,000 dilution of secondary antibody (sheep anti-mouse IgG conjugated with horseradish peroxidase; GE Healthcare) in Immuno Enhancer Solution 2 (Toyobo), and washed three times with TBST over 30 min. Signals were developed with the Western Lightning Chemiluminescence Reagent Plus (Perkin Elmer) and digital images acquired with a LAS-3000 imager (Fujifilm). Quantification of band intensities was performed using custom code written in Mathematica (Supplementary Information; Wolfram Research).

RNA/ChIP sequencing. Total RNA was extracted from semi-confluent cells grown in a 10-cm dish using TRIzol (Life Technologies). RNA sequencing was performed as described previously[39]. Complementary DNA was sequenced using the HiSeq 1500 system (Illumina). To calculate the total amount of each mRNA transcript, a series of programs, TopHat[40] (v1.4.1), and Cufflinks[41] (v1.3.0) were used. ChIP was performed as described previously[42]. Specifically, 2 µg of mouse antibody with 80 µl of anti-mouse IgG dynabeads (Life Technologies) were used for chromatin prepared from cells in one half of a semi-confluent 10-cm dish. After sonication (SLPE 40, Branson; 16% power, 8 min total, 55 s on and 5 s off), the median size of fragmented DNA was ~150 base pairs with a range of 50–300 base pairs. ChIPed DNA was sequenced using the HiSeq 1500 system (Illumina). The reads were aligned to the mouse (mm9) and MMTV genome using Bowtie software (version 0.12.8) with the following parameters: -v 3, -m 1. The FPKM (fragments per kilobase of transcript per million mapped reads) value of the sequenced reads was calculated every 10,000-base-pair bin with a shifting size of 1,000 base pairs. Additionally, the read number of the immunoprecipitated sample was normalized by subtracting the FPKM value of the input sample in each bin. Sequencing data was ranked and binned using custom written bash files and standard linux commands. Alignment of ChIP-seq data to binned sets of genes was done using the aggregation and correlation toolbox ACT[43]. All subsequent analysis of aggregated data, including plotting, was done using Mathematica (Wolfram Research). According to ChIP-seq data, RNAP2 was distributed symmetrically about transcription start sites (± 3 kilobases; Fig. 4b), consistent with the large percentage of active promoters that display divergent initiation[44].

Mathematical modelling and data fitting. The transcription cycle was modelled[15] as in Fig. 2d. The model can be written as a system of coupled first-order ordinary differential equations:

$$
\begin{aligned}
\frac{d}{dt}\text{Prom}(t) &= k_{in}(t) - (k_{out} + k_{ini})\text{Prom}(t) \\
\frac{d}{dt}\text{Init}(t) &= k_{ini}\text{Prom}(t) - (k_{esc} + k_{ab})\text{Init}(t) \\
\frac{d}{dt}\text{Elong}(t) &= k_{esc}\text{Init}(t) - k_t\text{Elong}(t)
\end{aligned}
\tag{6}
$$

Here Prom(t) is the number of RNAP2 that are bound and uninitiated within the gene array (CTD), Init(t) is the number of RNAP2 that are initiated at promoters within the gene array (Ser 5ph), Elong(t) is the number of RNAP2 that have escaped

the promoter region and are elongating within the gene array (Ser 2ph), and t is time. The various ks are transition rates between the different states, Prom, Init and Elong. k_{in}/k_{out} are the RNAP2 binding on/off rates, k_{ini}/k_{ab} are the RNAP2 initiation/abortion rates and k_{esc}/k_t are the RNAP2 promoter-escape/elongation-termination rates. For the initial conditions we assume that:

$$\text{Prom}(0) = \text{Init}(0) = \text{Elong}(0) = 0 \tag{7}$$

This kinetic model has been used previously to fit steady-state FRAP data such that two conditions are met: (1) the number of available RNAP2 binding sites does not change with time; and (2) diffusion is so fast that diffusive gradients have already equilibrated by the time the first data point in the FRAP recovery curve is acquired[15]. We made sure the first condition was met by performing FRAP only near the peak of transcriptional activation at the gene array, approximately 15 min after the addition of gene-activating hormone (dexamethasone). At this time RNAP2 levels at the gene array are near their peak (as seen in Fig. 1c) and remain fairly steady for around 15 min and beyond. We made sure the second condition was met by measuring the radial profile of our FRAP recovery over time. As we show in Extended Data Fig. 9c the radial profile does not change shape considerably after the first time point was collected, indicating diffusion can safely be neglected beyond this point. For this reason, we ignored the first acquired point from FRAP recovery curves when fitting (thereby fitting times from 1.4 s post-bleach onwards). With these conditions met, the RNAP2 binding on rate is time-independent, that is, $k_{in}(t) = C$, where C is an arbitrary constant. Substituting this into equation (6) makes the system linear and the analytic solution turns out to be independent of C. The FRAP recovery curve $FRAP(t)$ is proportional to Prom(t) + Init(t) + Elong(t). Renormalizing by the equilibrium condition (found by setting the left-hand side of equation (6) to zero) gives a FRAP recovery curve that scales from 0 to 1:

$$
\begin{aligned}
FRAP(t) &= \frac{(k_{ab}+k_{esc})k_t}{k_{esc}k_{ini}+(k_{ab}+k_{esc}+k_{ini})k_t} \\
&\left\{ \frac{e^{-(k_{ab}+k_{esc})t}k_{ini}(k_{ini}+k_{out})}{(k_{ab}+k_{esc}-k_{ini}-k_{out})} + \frac{e^{-(k_{ini}+k_{out})t}(k_{ab}+k_{esc}-k_{out})}{(k_{ini}+k_{out}-k_{ab}-k_{esc})} \right. \\
&+ \frac{e^{-(k_{ab}+k_{esc})t}k_{esc}k_{ini}(k_{ini}+k_{out})}{(k_{ab}+k_{esc})(k_{ini}+k_{out}-k_{ab}-k_{esc})(k_{ab}+k_{esc}-k_t)} \\
&- \frac{e^{-(k_{ini}+k_{out})t}k_{esc}k_{ini}}{(k_{ini}+k_{out}-k_{ab}-k_{esc})(k_{ini}+k_{out}-k_t)} \\
&- \frac{e^{-k_t t}k_{esc}k_{ini}(k_{ini}+k_{out})}{(k_{ab}+k_{esc}-k_t)(k_{ini}+k_{out}-k_t)k_t} \\
&\left. + \frac{k_{esc}k_{ini}+(k_{ab}+k_{esc}+k_{ini})k_t}{(k_{ab}+k_{esc})k_t} \right\}
\end{aligned}
\tag{8}
$$

FRAP data were renormalized (from 0 to 1) and fit to equation (8) using the built-in Mathematica function NonlinearModelFit (Wolfram Research). Data points were exponentially distributed in time to spread them more uniformly along fitted FRAP curves. This ensured the early part of the recovery was represented as much as the latter part, as is commonly done[34]. Each data point was weighted when fitting by the square of the standard deviation of the total FRAP data set divided by the square of the experimental error of that data point.

For fitting Fab recruitment data, we extended the model to allow the number of available binding sites for RNAP2 to change with time, as would be expected upon gene activation. Specifically, we let

$$k_{in}(t) = k_{in}^{max}\text{GR}(t-\Delta)[n_{sites}-\text{Prom}(t)-\text{Init}(t)] \tag{9}$$

Where n_{sites} is the total number of RNAP2 binding sites within the gene array. This assumes that the sites available to freely-diffusing RNAP2 are proportional to the number of sites available to GR. Here GR(t) is a phenomenological function that fits the average GR recruitment curve rescaled from 0 to 1:

$$\text{GR}(t) \equiv \frac{a}{1+e^{-k_1(t-t_0)}}e^{-k_2(t-t_0)} \tag{10}$$

with $a = 1.22$, $k_1 = 0.50$, $k_2 = 0.024$ and $t_0 = 10$ as starting estimates whose final values were determined from fits to single cell data (see Extended Data Fig. 9a for a sample fit). Δ is the time delay between a promoter site being available for GR and it being available for RNAP2. Δ could be due to either necessary chromatin remodelling or necessary recruitment of other factors in a pre-initiation complex, or some combination of these. n_S is the total number of available promoter-proximal RNAP2 sites within the array. With this model, the binding on rate $k_{in}(t)$ can go up/down in two ways: either a promoter-proximal site is made accessible/inaccessible to GR

(the GR(t) term) or a promoter-proximal site is filled/unfilled by another RNAP2 (the $n_{sites} - \text{Prom}(t) - \text{Init}(t)$ term).

Because GR(t) is time-dependent, equation (6) is nonlinear, so an analytic solution is not available. Fab recruitment curves were therefore fitted to a numerical solution to equation (6) (with equation (9) in place) obtained with the built-in Mathematica function NDSolve (Wolfram Research). Fitting to the numerical solution was then done with the built-in Mathematica function NonlinearModelFit (Wolfram Research) using the same error-based weighting scheme as described for FRAP data above. CTD Fab recruitment curves were fit to the numerical solution for $\text{Prom}(t) + \text{Init}(t) + \text{Elong}(t)$, Ser 5ph Fab recruitment curves were fit to the numerical solution for $\text{Init}(t) + \text{Elong}(t)$, and Ser 2ph Fab recruitment curves were fit to the numerical solution for $\text{Elong}(t)$.

To perform simultaneous fits of FRAP and Fab data, each data point was assigned an extra dimension whose value could be $i = 1, 2, 3, 4$ corresponding to which curve it belonged to, either the CTD, Ser 5ph, or Ser 2ph Fab recruitment curve, or the FRAP recovery curve, respectively. Fits of this new data set were done as before using the following two-dimensional function:

$$f(i,t) = \begin{cases} \text{Prom}(t) + \text{Init}(t) + \text{Elong}(t) & , i = 1 \\ \text{Init}(t) + \text{Elong}(t) & , i = 2 \\ \text{Elong}(t) & , i = 3 \\ \text{FRAP}(t) & , i = 4 \end{cases} \quad (11)$$

26. Kimura, H., Hayashi-Takanaka, Y. & Yamagata, K. Visualization of DNA methylation and histone modifications in living cells. *Curr. Opin. Cell Biol.* **22,** 412–418 (2010).
27. Manders, E. M., Kimura, H. & Cook, P. R. Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J. Cell Biol.* **144,** 813–822 (1999).
28. McNeil, P. L. & Warder, E. Glass beads load macromolecules into living cells. *J. Cell Sci.* **88,** 669–678 (1987).
29. Sugaya, K., Vigneron, M. & Cook, P. R. Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein. *J. Cell Sci.* **113,** 2679–2683 (2000).
30. Gronostajski, R. M. Roles of the NFI/CTF gene family in transcription and development. *Gene* **249,** 31–45 (2000).
31. Mueller, F., Morisaki, T., Mazza, D. & McNally, J. G. Minimizing the impact of photoswitching of fluorescent proteins on FRAP analysis. *Biophys. J.* **102,** 1656–1665 (2012).
32. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8,** 377–385 (1998).
33. Mueller, F., Wach, P. & McNally, J. G. Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. *Biophys. J.* **94,** 3323–3339 (2008).
34. Stasevich, T. J., Mueller, F., Brown, D. T. & McNally, J. G. Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. *EMBO J.* **29,** 1225–1234 (2010).
35. Kimura, H., Tao, Y., Roeder, R. G. & Cook, P. R. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Mol. Cell. Biol.* **19,** 5383–5392 (1999).
36. Morris, D. P., Michelotti, G. A. & Schwinn, D. A. Evidence that phosphorylation of the RNA polymerase II carboxyl-terminal repeats is similar in yeast and humans. *J. Biol. Chem.* **280,** 31368–31377 (2005).
37. Li, M. *et al.* Solution structure of the Set2–Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc. Natl Acad. Sci. USA* **102,** 17636–17641 (2005).
38. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20,** 2922–2936 (2006).
39. Odawara, J. *et al.* The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. *BMC Genomics* **12,** 516 (2011).
40. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
41. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
42. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33,** 61–73 (2008).
43. Jee, J. *et al.* ACT: aggregation and correlation toolbox for analyses of genome tracks. *Bioinformatics* **27,** 1152–1154 (2011).
44. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).
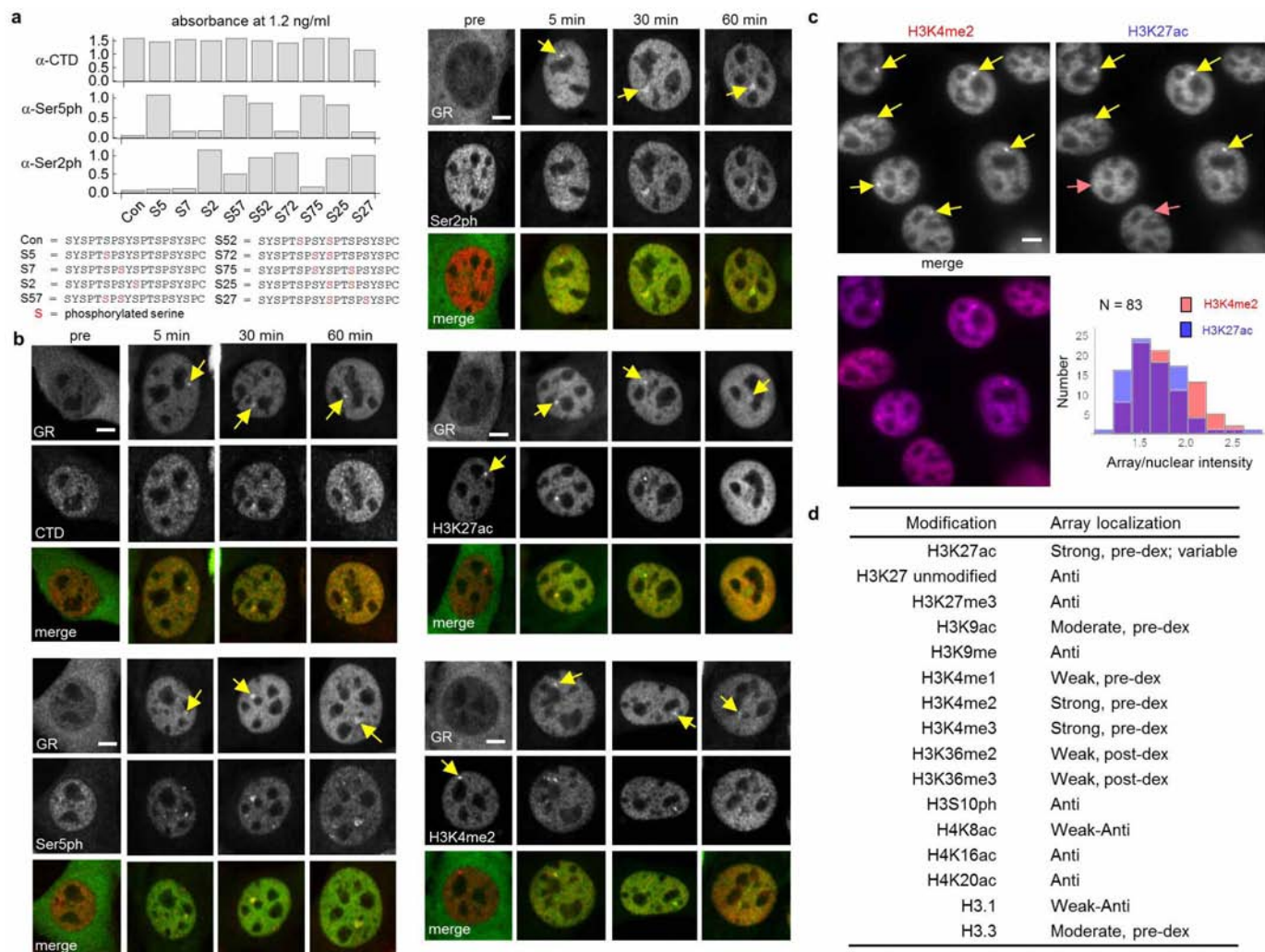
Extended Data Figure 1 | Immunostaining the array with antibodies against RNAP2 and histones. a, Monoclonal antibodies against the RPB1 subunit of RNAP2 (CTD) and its Ser 5 and Ser 2 phosphorylated forms (Ser 5ph and Ser 2ph) were evaluated by ELISA using the indicated peptides with specific phosphorylation patterns. Microtitre plates coated with the peptides were incubated with each antibody. After incubation with peroxidase-conjugated secondary antibody and washing, the colorimetric signal of tetramethylbenzidine was detected by measuring the absorbance at 405 nm using a plate reader. **b**, Immunofluorescence with monoclonal antibodies against RNAP2 (tested in **a**), H3K27ac and H3K4me2 in cells (arrays marked by GR, yellow arrows) fixed pre- and post-transcriptional activation (times indicated). **c**, Although H3K4me2 levels at the array are consistently high, H3K27ac levels are sometimes relatively low (pink arrows), as quantified in the histogram. **d**, Summary of a screen of histone modifications and variants at the MMTV array by immunostaining. Scale bars, 5 μm.
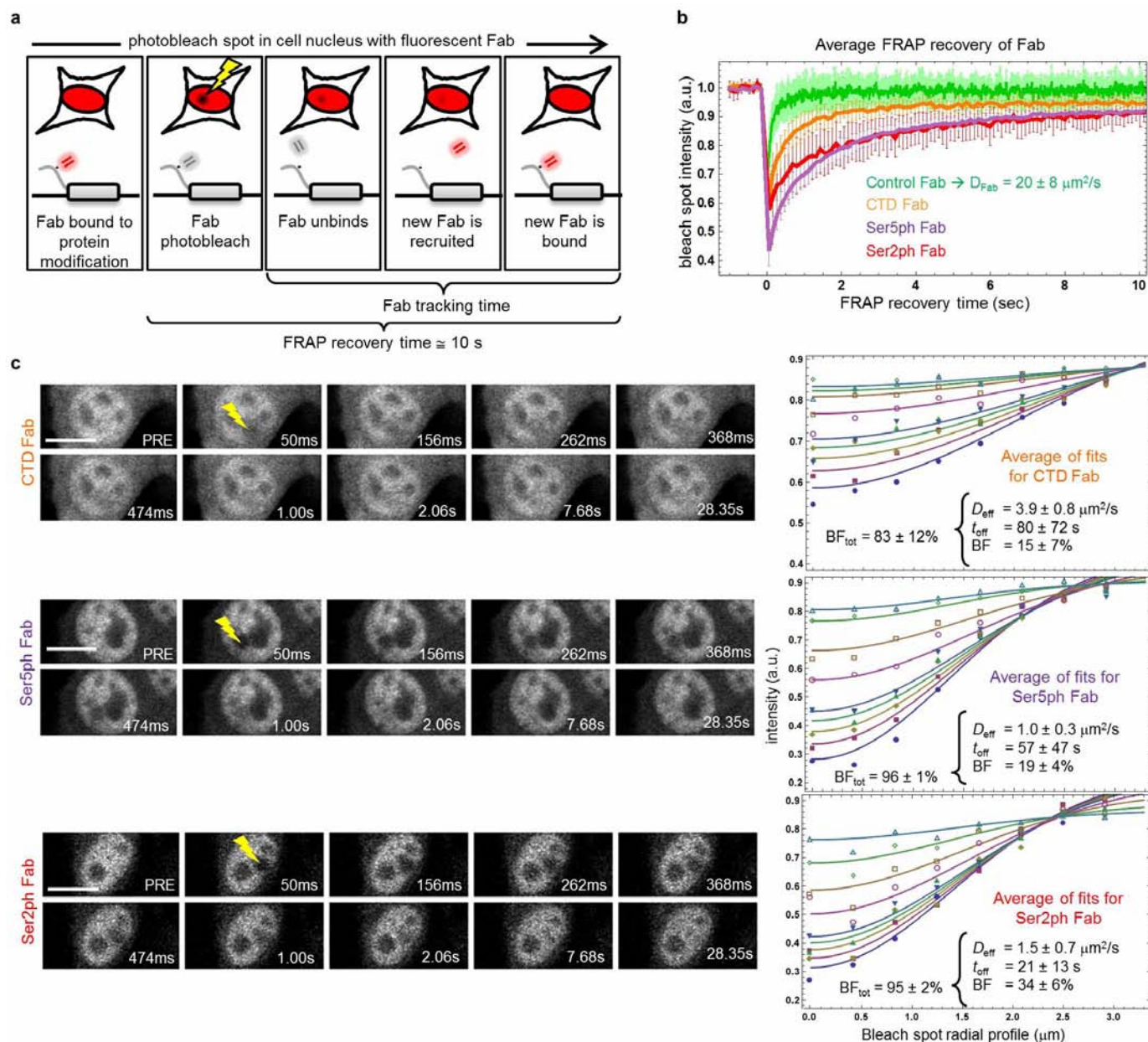
Extended Data Figure 2 | FRAP determines the timescale of FabLEM experiments. a, FRAP experiments on cells loaded with fluorescent Fabs (red) against RNAP2 (CTD, Ser 5ph and Ser 2ph) can be used to estimate an upper bound on how long it takes Fabs to track their targets. The FRAP recovery time is limited by the dissociation of photobleached Fab (grey) from target protein modifications (black dots) plus the association time of an unbleached Fab to the open modification. It is the latter time that corresponds to the tracking time of Fab and the temporal resolution of FabLEM. **b**, Fab FRAP recovery curves (coloured curves on right; ± s.e.m.) are complete in about 10 s. A control Fab with no target in the nucleus is shown in green to see how fast recoveries are when Fab do not bind (fits to $n = 10$ curves indicate this recovery is purely diffusive, yielding a Fab diffusion coefficient ± s.e.m. of $D_{Fab} = 20 \pm 8\ \mu m^2\ sec^{-1}$). **c**, Select frames from single-cell FRAP experiments. Yellow lightening indicates the position of the bleach. On the right, reaction-diffusion fits of the bleach spot profiles (radial distance from centre of bleach spot) at times shown on the left, with earlier times having deeper profiles. Fits (± s.e.m.) to $n = 43$ (CTD), $n = 30$ (Ser 5ph), and $n = 32$ (Ser 2ph) FRAP experiments on three independent days yield the average effective diffusion coefficient D_{eff}, binding time t_{off}, and bound fraction BF for each Fab. The total bound fraction (BF$_{tot}$) is computed: $1 - (D_{eff}/D_{Fab})(1 - BF)$. All Fab have high total bound fractions (>80%), indicating good signal to noise. Scale bars, 10 μm; a.u., arbitrary units.
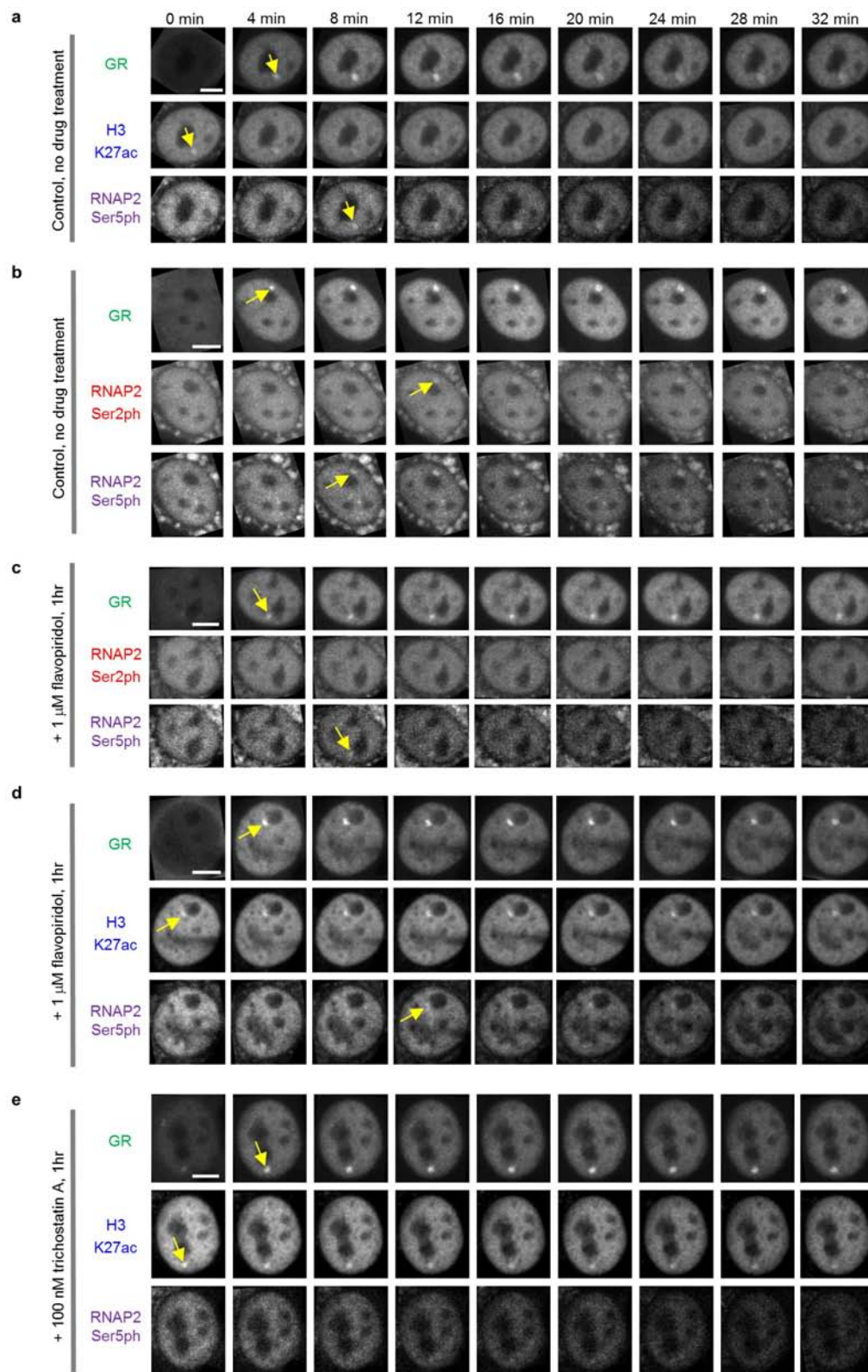
Extended Data Figure 3 | Testing fidelity of FabLEM for detecting RNAP2 and histone modifications. a–e, Representative FabLEM experiments in single living cells that were either untreated (**a**, **b**) or treated with inhibitors (**c–e**). Scale bars 5 μm. **c,** Cells were treated with 1 μM flavopiridol for 1 h and then activated with hormone (time roughly corresponds to post-activation). Flavopiridol inhibits P-TEFb, which phosphorylates the RNAP2 CTD at Ser 2, preventing elongation at the gene array (marked by yellow arrow). FabLEM experiments confirm this, showing no array decondensation and no accumulation of Ser 2ph Fab (red) at the array upon hormone treatment even though Ser 5ph (purple) does accumulate, indicating RNAP2 initiation (although to a lesser extent than in untreated cells). **d,** The same experiment as above, but now examining histone acetylation levels at the array (blue, H3K27ac), which no longer go down post-activation, as in untreated cells. **e,** Cells were treated with 100 nM of the histone deacetylase inhibitor trichostatin A (TSA) for 1 h. As with flavopiridol treatment, the array no longer decondenses, H3K27ac levels remain high (and in fact the total nuclear intensity is higher, indicating global increases in H3K27ac levels) and levels of RNAP2 initiation are low, indicating little or no RNAP2.
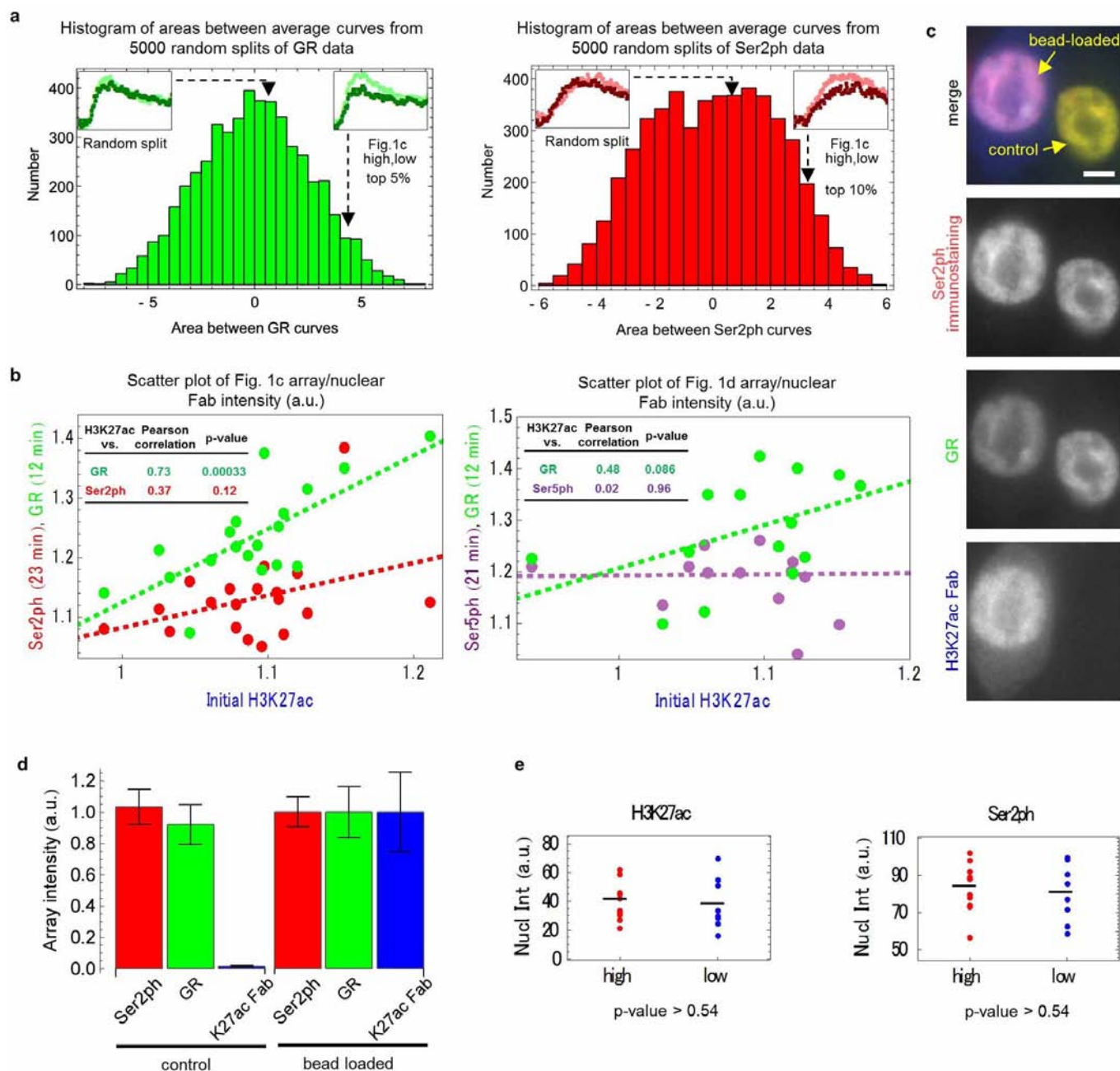
Extended Data Figure 4 | Testing the confidence of FabLEM correlations.
a, GR and Ser 2ph data from Fig. 1c were randomly split into two groups in 5,000 unique ways and the area between the average curves from each group was computed. A histogram of the obtained areas is shown. This reveals that the H3K27ac-based split of data into high and low groups in Fig. 1c scores in the top 5% of all GR splits (green) and in the top 10% of all Ser 2ph splits (red). This provides an estimate for the confidence of the measured correlation between H3K27ac and GR or Ser 2ph. **b**, Scatter plots of single-cell data from Fig. 1c, d with the initial H3K27ac array/nuclear intensity (2 ± 2 min) on the *x*-axis and the maximal GR (12 ± 2 min, green), Ser 2ph (23 ± 2 min, red) and Ser 5ph (21 ± 2 min, purple) array/nuclear intensities on the *y*-axis. Each point represents data from a single cell averaged over a four-minute time window (for example, each Ser 2ph point represents the mean of data from a single cell between 21 and 25 min). A positive correlation (quantified by the Pearson correlation coefficient and its corresponding *P* value calculated using the built-in Mathematica function PearsonCorrelationTest; Wolfram Research) is seen between H3K27ac and GR, and H3K27ac and Ser 2ph, but not between H3K27ac and Ser 5ph. **c**, To test if the nuclear concentration of loaded Fab has no deleterious effect on transcription at the array and is not responsible for the H3K27ac-dependence of GR recruitment and RNAP2 elongation (Ser 2ph) shown in Fig. 1c, immunostaining against Ser 2ph (red) was performed on a population of induced cells (30 min) expressing GFP–GR (green) in which a subset were bead loaded with the H3K27ac-specific Fab (blue). **d**, The intensity of arrays in control unloaded cells ($n = 24$) was the same within error (± s.e.m.) as in bead-loaded cells ($n = 24$), indicating the H3K27ac Fab do not alter Ser 2ph levels at the array. **e**, The Fab nuclear intensities of high/low sorted cells based on array intensities from Fig. 1c are statistically indistinguishable (the smallest *P* value from the Student's *t*-test, the *z*-test and the Mann–Whitney median test is reported using the built-in Mathematica function LocationTest; Wolfram Research). This demonstrates that differing concentrations of Fab in the nucleus are not responsible for the measured correlation between H3K27ac and Ser 2ph.
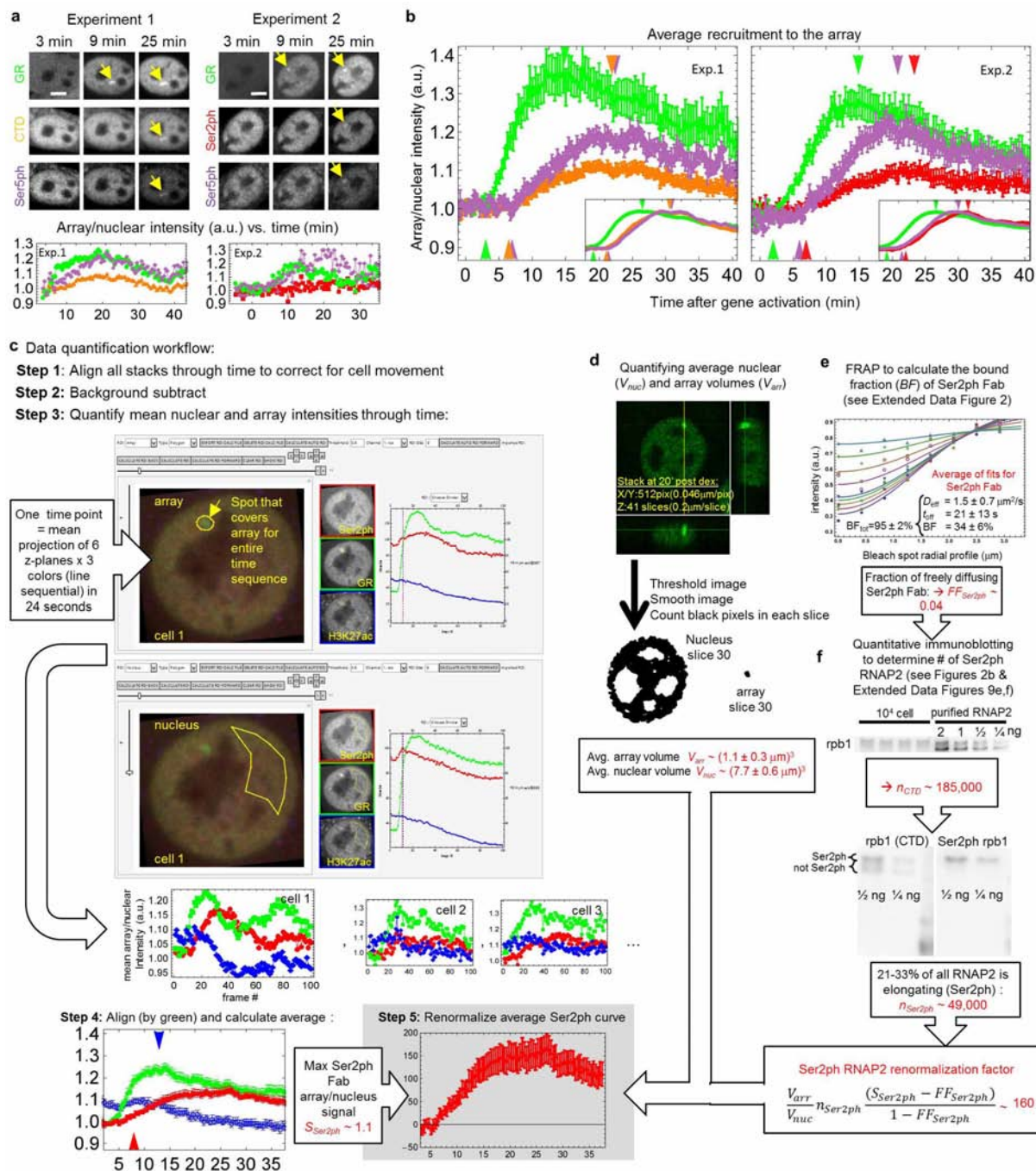
Extended Data Figure 5 | FabLEM quantification. a, Frames from sample single-cell experiments show how the gene array (yellow arrow) is first bound by GFP–GR (green) after hormone is added to cells, followed by Fabs marking RNAP2 (orange, CTD, recruitment), Ser 5 phosphorylated RNAP2 (blue, Ser 5ph, initiation), and Ser 2 phosphorylated RNAP2 (red, Ser 2ph, elongation). Scale bars, 5 μm. Quantification of the array/nuclear intensity over time is shown below after adjusting time scale so GR curves are aligned (see Extended Data Fig. 9a for alignment details). **b**, Average single-cell recruitment curves ($n = 12$, ± s.e.m.). Insets show a normalized rolling average to illustrate the temporal ordering: gene activation, RNAP2 recruitment, initiation, and elongation. The arrows indicate, from left to right, when levels of GR, RNAP2 and RNAP2 Ser 5ph and Ser 2ph go up/down at the array. **c–f**, Summary of workflow for quantifying FabLEM data. Image stacks were aligned in time (step 1), background subtracted (step 2), and intensities in the red, green, and blue channels were measured in regions of interest covering both the array (yellow polygon in upper screen shot) and a representative portion of the nucleus (yellow polygon in lower screen shot). From this raw intensity data, the mean array/nuclear intensity was calculated for the cell. This was repeated for

other cells and data averaged by aligning green GFP–GR curves (step 4). The maximum array relative to nuclear Ser 2ph signal is $S_{Ser2ph} \approx 1.1$. **d**, The average volume of the array V_{arr} and nucleus V_{nuc} were calculated from image stacks of cells expressing GFP–GR twenty minutes after transcription activation by dexamethasone. Image stacks were smoothed with a median filter (to remove single voxel speckle noise) and binarized by making all voxels with intensities above a threshold value black and all voxels equal to or below the threshold value white. The volume of the nucleus and the array could then be estimated by counting the number of black voxels. **e**, FRAP experiments on cells loaded with fluorescent Ser 2ph Fab were performed and fit with a reaction-diffusion model to estimate the total bound fraction of Ser 2ph Fab (BF_{tot}) from which the free fraction could be calculated $FF_{Ser2ph} = 1 - BF_{tot} \approx 0.04$. **f**, Quantitative immunoblotting was used to estimate the total number of RNAP2 per cell, n_{CTD}, as well as the number phosphorylated at Ser 2, n_{Ser2ph}. Together the estimates of S_{Ser2ph}, V_{arr}, V_{nuc}, FF_{Ser2ph} and n_{Ser2ph} from **c–f** were used to calculate the Ser 2ph renormalization factor and generate the final renormalized FabLEM Ser 2ph curve (step 5). The renormalized curves in Fig. 2a were generated in an analogous manner.
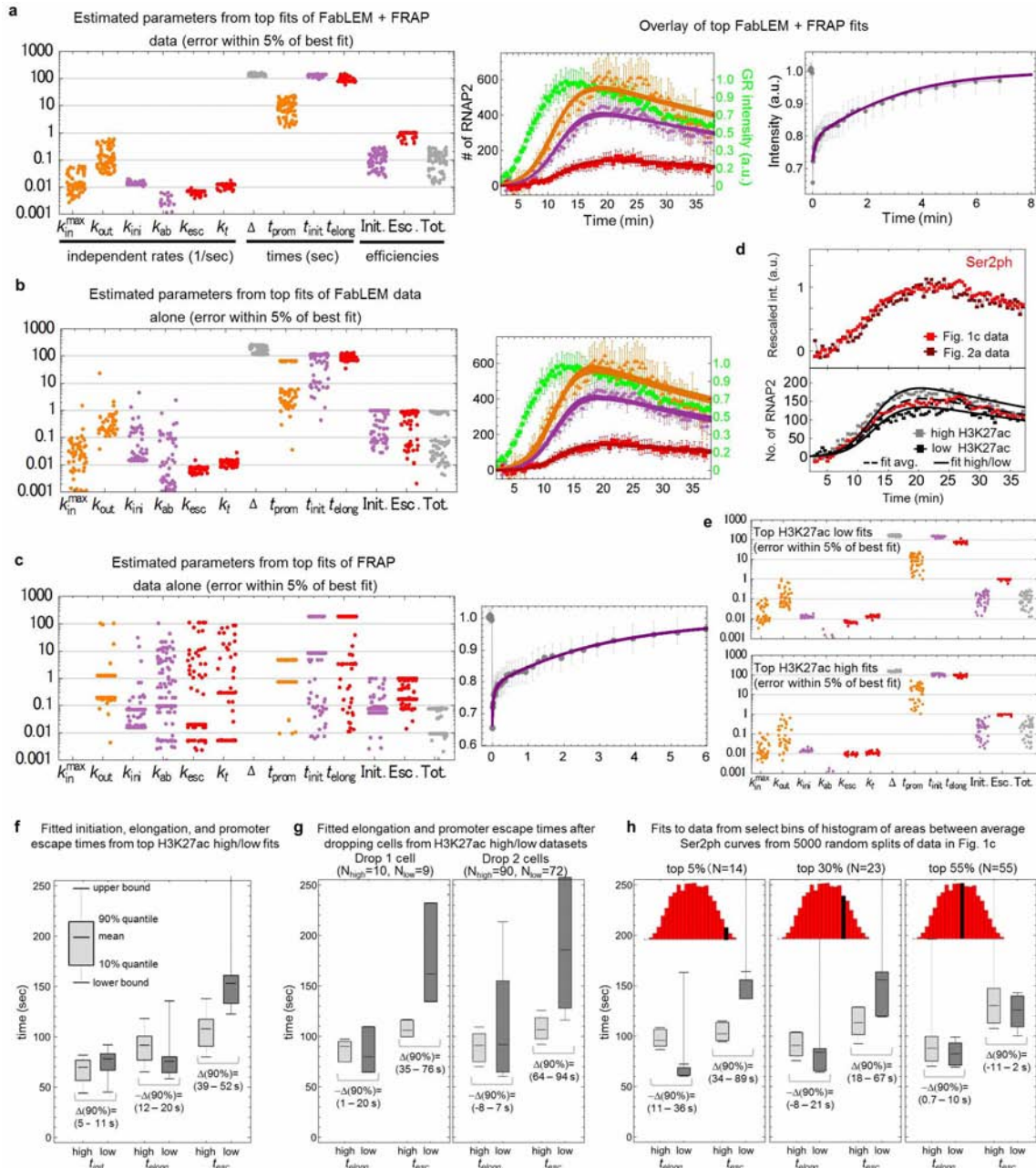
Extended Data Figure 6 | Testing the quality of FabLEM fits. a, When FRAP and FabLEM data are simultaneously fit to the same model, the top fits (within 5% error of the best fit) found after searching the full parameter space are better constrained compared to fits of only FabLEM data in **b** or fits of only FRAP data in **c**, so all parameters can be estimated within an order of magnitude. Note that FRAP experiments are performed in steady-state, so fits do not depend on k_{in}^{max} or Δ. **d**, Ser 2ph data from experiments using Fab against H3K27ac (Fig. 1c) are consistent with data from experiments using Fab against Ser 5ph (Fig. 2a and Extended Data Figs 5a, b). This indicates the sorted cell Ser 2ph data from Fig. 1c can be used in place of the Ser 2ph data in Fig. 2a for fitting purposes, as shown in the lower panel. **e**, Fits to data taken from sorted cells with arrays having low (upper panel) or high (lower panel) H3K27ac levels before transcriptional activation. The three parameters ($t_{init} = 1/k_{ini}$, $t_{elong} = 1/k_t$, and $t_{esc} = 1/k_{esc}$) that change the most significantly between these fits are plotted for comparison in **f**. The mean, the 10/90% quantiles, and data bounds are shown. If the 10/90% quantiles do not overlap, then the fitted parameters are statistically different with >99% confidence (that is, the top 10% of the top 10%). Of these parameters, only t_{esc} did not have overlapping 10/90% quantiles. The 90% mean difference confidence interval Δ(90%) for the high/low fitted parameters is reported (calculated with the built-in Mathematica function MeanDifferenceCI; Wolfram Research). **g**, To confirm the statistical

significance of the high/low H3K27ac fits, one or two random cells were dropped from each high ($N = 9$) and low ($N = 10$) group in all possible ways (N_{high} and N_{low} denoted). The average curves of these subgroups were fit and results for t_{esc} and t_{elong} are shown. In all cases, the 10/90% quantiles for t_{esc} do not overlap between high/low groups. In contrast, the 10/90% quantiles for t_{elong} do overlap, meaning there is less statistical difference between t_{elong} in high/low H3K27ac cells. **h**, To further cross-validate, select random splits of the data from a histogram like the one in Extended Data Fig. 4a were fit in the same way as **f** and **g**. The black area of the histogram (inset) shows which bin splits were taken from for fitting (N splits from each bin). A comparison of **f** and **h** reveals that even though the H3K27ac sorted high/low split only ranks in the top 10% of random splits (ranked by area between split average curves, see Extended Data Fig. 4a), fitted t_{esc} values in **f** and **g** between high/low groups are statistically as distinct as those in **h** from the top 5% of random splits. In contrast, fitted t_{elong} values are not statistically distinct in **f** and **g**, but they are in **h**. This demonstrates that sorting data by the initial levels of H3K27ac (like in **f** and **g**) is better at distinguishing fast t_{esc} from slow t_{elong}, whereas random data sorted solely by the area between split curves (like in **h**) cannot distinguish these two effects. Thus, the difference between fitted t_{esc} is statistically significant and supports a link between H3K27ac and the RNAP2 promoter escape rate rather than between H3K27ac and the RNAP2 elongation rate.
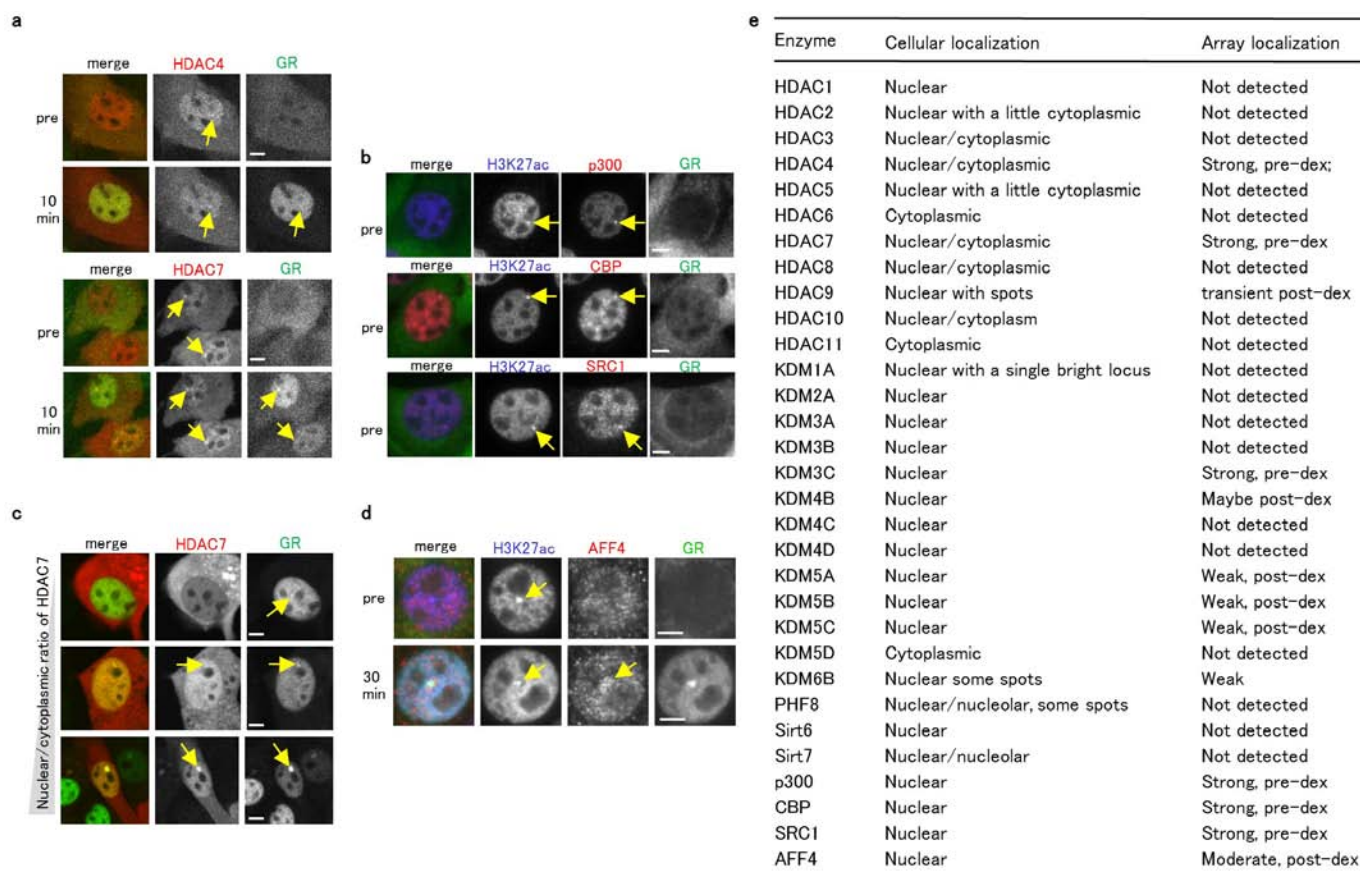
Extended Data Figure 7 | Screening the MMTV array for histone-modifying enzymes. a, The histone deacetylases HDAC4 (HaloTag, top, red) and HDAC7 (HaloTag, bottom, red) colocalize with the array (yellow arrows) both before (pre) and after (10 min) transcription activation by GR (green). **b**, The lysine acetyltransferases (KATs) p300 (top, red) and CBP (middle, red) also colocalize with the array pre-activation (as marked by H3K27ac, blue), as does the steroid receptor cofactor 1 (SRC1, bottom, red), an adaptor that bridges p300/CBP to GR after activation. **c**, HDAC7 (and HDAC4, data not shown) is distributed both in the cytoplasm and nucleus and the cytoplasmic/nuclear intensity ratio varies from cell to cell. When in the nucleus, they colocalize with the array (bottom two rows). **d**, AFF4 (red) does not colocalize with H3K27ac (blue) at arrays pre-activation. However, AFF4 can be seen 30 min after gene activation, along with GFP–GR (green) and H3K27ac (blue). **e**, A variety of histone-modifying enzymes were tested to see which localized at the array. All HDACs, KDMs, Sirts and also PHF8 were HaloTag-tagged and screened by transient transfection, while the remaining enzymes were screened by immunostaining. Scale bars 5 μm.

Enzyme	Cellular localization	Array localization
HDAC1	Nuclear	Not detected
HDAC2	Nuclear with a little cytoplasmic	Not detected
HDAC3	Nuclear/cytoplasmic	Not detected
HDAC4	Nuclear/cytoplasmic	Strong, pre-dex;
HDAC5	Nuclear with a little cytoplasmic	Not detected
HDAC6	Cytoplasmic	Not detected
HDAC7	Nuclear/cytoplasmic	Strong, pre-dex
HDAC8	Nuclear/cytoplasmic	Not detected
HDAC9	Nuclear with spots	transient post-dex
HDAC10	Nuclear/cytoplasm	Not detected
HDAC11	Cytoplasmic	Not detected
KDM1A	Nuclear with a single bright locus	Not detected
KDM2A	Nuclear	Not detected
KDM3A	Nuclear	Not detected
KDM3B	Nuclear	Not detected
KDM3C	Nuclear	Strong, pre-dex
KDM4B	Nuclear	Maybe post-dex
KDM4C	Nuclear	Not detected
KDM4D	Nuclear	Not detected
KDM5A	Nuclear	Weak, post-dex
KDM5B	Nuclear	Weak, post-dex
KDM5C	Nuclear	Weak, post-dex
KDM5D	Cytoplasmic	Not detected
KDM6B	Nuclear some spots	Weak
PHF8	Nuclear/nucleolar, some spots	Not detected
Sirt6	Nuclear	Not detected
Sirt7	Nuclear/nucleolar	Not detected
p300	Nuclear	Strong, pre-dex
CBP	Nuclear	Strong, pre-dex
SRC1	Nuclear	Strong, pre-dex
AFF4	Nuclear	Moderate, post-dex

Extended Data Figure 8 | The ratio of elongating to promoter-bound RNAP2 increases with H3K27ac levels independent of RNA expression level. a, Histograms of H3K27ac levels above input at genes (transcription start site TSS ± 2,000 base pairs) with varying levels of activity according to RNA sequencing. Each plot corresponds to 1,000 genes, with the top 1,000 most active genes in the very top plot followed successively underneath by plots showing the 1,000–2,000, 2,000–3,000…7,000–8,000 most active genes. A trend can be seen, with the more active genes having on average slightly higher acetylation than less active genes, although there is huge variability in all bins suggesting some very active genes have little H3K27ac, while other inactive genes have lots of H3K27ac. **b**, When RNAP2 occupancy is mapped to these binned genes (with colours corresponding to the histograms), those with higher H3K27ac levels tend to have more elongating RNAP2 relative to the amount bound at the promoter (± 350 base pairs, easiest to see in the renormalized plots on the far right), regardless of RNA expression level.
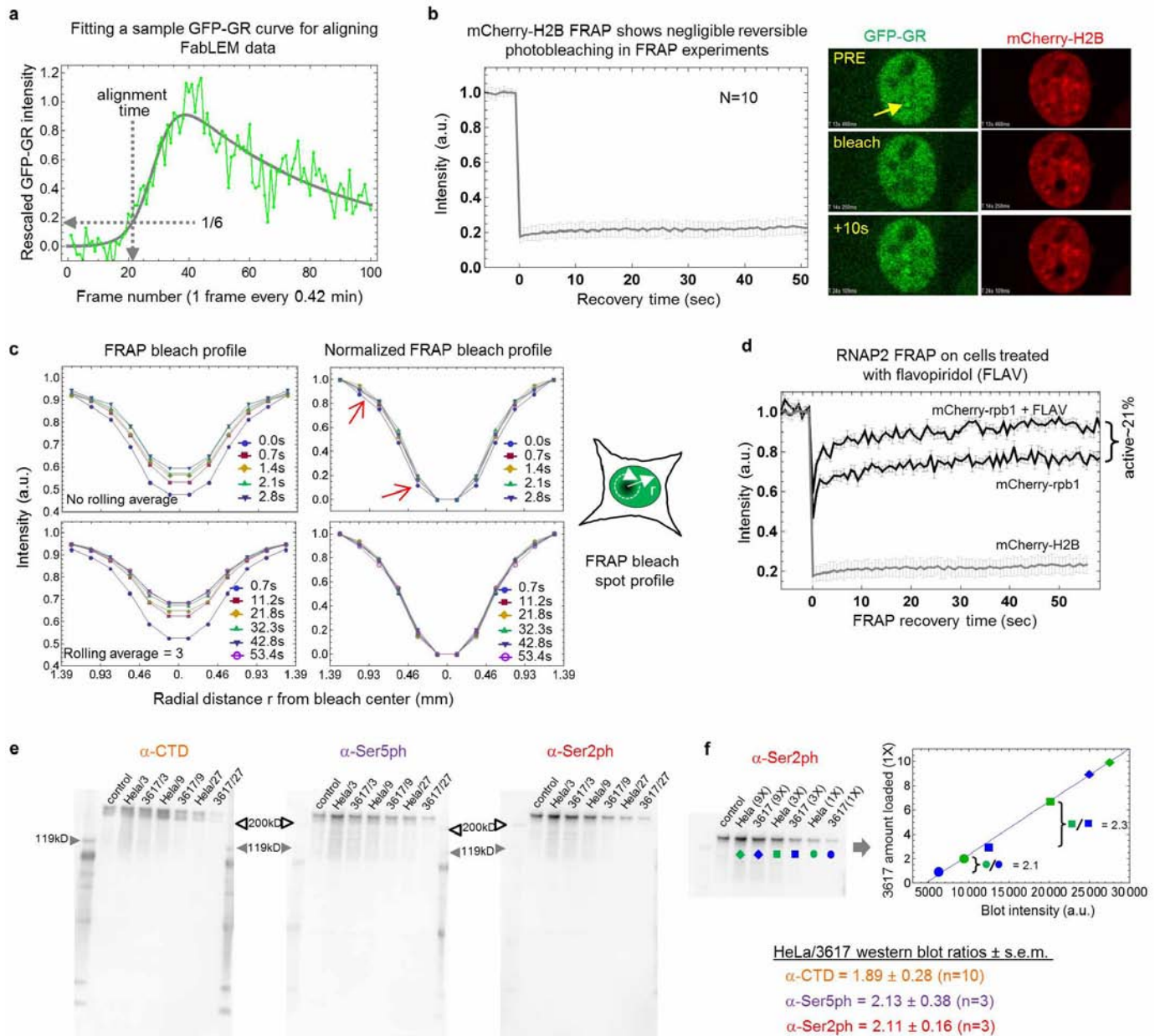
Extended Data Figure 9 | FabLEM data alignment, FRAP checks and immunoblot quantification. **a**, The GFP–GR curve is used to align FabLEM curves by fitting to a phenomenological model describing the basic structure of the average GFP–GR curve rescaled from 0 to 1. The fitting curve has the general form $a \exp(-k_2[t-t_0])/(1-\exp[-k_1(t-t_0)])$ with a, k_1, k_2 and t_0 left as fitting parameters whose starting estimated values were determined by fitting the average of all GFP–GR recruitment curves. After fitting an individual curve to this function, we numerically determine when the fitted function has a value of ~0.16. This time is used as the aligning time for the individual curve, which we define as 4.2 min post-activation (corresponding to frame 10 post-activation). **b**, To ensure reversible photobleaching was minimal in our FRAP experiments on the mCherry-labelled RPB1 subunit of RNAP2, mCherry–RPB1, we duplicated experiments on histone H2B (mCherry–H2B, $n = 10$). This showed very little fluorescence recovery 50 s post-bleach at the array (yellow arrow) where GFP–GR was colocalized ($\pm$ s.e.m.), demonstrating negligible reversible photobleaching of mCherry. **c**, To check for the role of diffusion in mCherry–RPB1 recoveries after photobleaching at the gene array (see Fig. 2c for an example), the profile of the photobleach spot was measured with time (defined in cartoon to right). Top row: after normalizing from 0 to 1, the curves all fall on top of each other, indicating no diffusive recovery (which

drives spatial distortions in the bleach spot shape). Only the first post-bleach time point (labelled $t = 0.0$ s here) shows a difference in shape (red arrows) from the others (0.7 s, 1.4 s,...), indicating diffusion plays a role in the recovery up until about 0.7 s. Bottom row: this is further confirmed by doing a rolling average of data (averaging three frames at a time), which shows no distortion of shape after normalization all the way up to 54 s post-bleach. **d**, FRAP was performed in cells transiently expressing mCherry–RPB1 and treated ($n = 27$) or untreated ($n = 27$) with the elongation inhibiting drug flavopiridol (1 μM, 1 h). In treated cells, the recovery was 21% more complete at 50 s post-bleach (relative to the baseline recovery from mCherry–H2B FRAP in **b**) than in untreated cells, indicating this fraction of RNAP2 is elongating in the untreated cells. **e**, Sample immunoblots of whole-cell extract from 3617 and HeLa cells using monoclonal antibodies against the CTD (α-CTD) of RPB1, as well as its Ser-5 (α-Ser 5ph) and Ser-2 (α-Ser 2ph) phosphorylated forms. **f**, Sample immunoblot quantification. Images were digitized and the intensity of blots quantified (blue/green shapes). Here the intensity of HeLa blots was fitted to a line to determine the relative intensity of blots from 3617 cells (within the linear region of the fit) and thereby determine the ratios of the numbers of RNAP2 in each cell type.

Extended Data Table 1 | Summary of estimated model parameters

Parameter	Estimated value (range)	Technique
No. of copies in MMTV gene array	175 (150-200)	DNA-Seq/McNally et al.
Length of single copy of MMTV array L_{arr}	9 kb (8-10)	DNA-Seq/McNally et al.
Length of MMTV array transcription unit L_{tr}	> 1 kb	DNA-Seq/McNally et al.
Array RNA (25 min post-Dex vs. 0, 5 min)	6.4X increase	RNA-Seq
Nuclear volume V_{nuc}	457 μm^3 (424 − 491)	microscopy
Array volume V_{arr}	1.33 μm^3 (1.02 − 1.71)	microscopy
No. of RNAP2/nucleus n_{CTD}	185,000 (165,000 − 205,000)	WB
No. of Ser5ph RNAP2/nucleus n_{Ser5ph}	123,000 (108,000 − 138,000)	WB
No. of Ser2ph RNAP2/nucleus n_{Ser2ph}	49,000 (42,000 − 56,000)	WB, FRAP
Max CTD Fab array/nucleus signal S_{CTD}	1.11 (1.09 − 1.13)	FabLEM
Max Ser5ph Fab array/nucleus signal S_{Ser5ph}	1.21 (1.18 − 1.24)	FabLEM
Max Ser2ph Fab array/nucleus signal S_{Ser2ph}	1.10 (1.08 − 1.12)	FabLEM
CTD Fab free fraction FF_{CTD}	0.17 (0.05 − 0.29)	FRAP
Ser5ph Fab free fraction FF_{Ser5ph}	0.04 (0.03 − 0.05)	FRAP
Ser2ph Fab free fraction FF_{Ser2ph}	0.05 (0.03 − 0.07)	FRAP
RNAP2 renormalization factor	610 (540 − 680)	$\frac{V_{arr}}{V_{nuc}} n_{CTD} \frac{(S_{CTD} − FF_{CTD})}{1 − FF_{CTD}}$
Ser5ph RNAP2 renormalization factor	440 (390 − 490)	$\frac{V_{arr}}{V_{nuc}} n_{Ser5ph} \frac{(S_{Ser5ph} − FF_{Ser5ph})}{1 − FF_{Ser5ph}}$
Ser2ph RNAP2 renormalization factor	160 (140 − 180)	$\frac{V_{arr}}{V_{nuc}} n_{Ser2ph} \frac{(S_{Ser2ph} − FF_{Ser2ph})}{1 − FF_{Ser2ph}}$
RNAP2 promoter sampling rate k_{in}^{max}	1.1/min (0.16 − 3.68)	FabLEM + FRAP fit
RNAP2 promoter unbinding rate k_{out}	9.3/min (2 − 35)	FabLEM + FRAP fit
RNAP2 promoter initiation rate k_{ini}	0.78/min (0.63 − 1.2)	FabLEM + FRAP fit
RNAP2 promoter abort rate k_{ab}	0.04/min (0 − 0.37)	FabLEM + FRAP fit
RNAP2 promoter escape rate k_{esc}	0.40/min (0.24 − 0.55)	FabLEM + FRAP fit
RNAP2 elongation rate k_t	0.70/min (0.40 − 1.0)	FabLEM + FRAP fit
RNAP2 recruitment time after GR Δ	2.3 min (1.9 − 2.9)	FabLEM + FRAP fit
Time RNAP2 is at promoter uninitiated	0.17 min (0.03 − 0.4)	$1/(k_{out} + k_{ini})$
Time RNAP2 is at promoter initiated	2.3 min (1.5 − 2.6)	$1/(k_{esc} + k_{ab})$
Time RNAP2 is elongating t_{elong}	1.4 min (1.0 − 2.5)	$1/k_t$
RNAP2 elongation rate	> 0.4 kb/min	$L_{tr} k_t$
RNAP2 initiation efficiency	13% (2 − 30)	$k_{ini}/(k_{ini} + k_{out})$
RNAP2 promoter escape efficiency	90% (41 − 100)	$k_{esc}/(k_{esc} + k_{ab})$
Slow down in elongation time in high vs. low H3K27ac cells Δt_{elong}	0.27 min (0.19 − 0.34; 90% mean difference confidence interval)	$\Delta(1/k_t)$ FabLEM + FRAP H3K27ac high/low fit
Speed up in promoter escape time in high vs. low H3K27ac cells Δt_{esc}	0.75 min (0.65 − 0.86; 90% mean difference confidence interval)	$\Delta(1/k_{esc})$ FabLEM + FRAP H3K27ac high/low fit

Each row shows the parameter name, the estimated value, and the technique used to estimate the value. In some cases parameters were independently estimated by McNally et al.[7] For fitted data, range corresponds to the 90% confidence interval of the best fit (or, approximately equivalent in this case, the range of estimated parameters from all fits whose error is within 5% of the best fit, see Extended Data Fig. 6). The 90% mean difference confidence interval is reported for Δt_{esc} and Δt_{elong}, as described in Extended Data Fig. 6f. For measured data, the range corresponds to the standard error of the mean ($\pm$ s.e.m.).

CORRECTIONS & AMENDMENTS

Corrigendum: Producing primate embryonic stem cells by somatic cell nuclear transfer

J. A. Byrne, D. A. Pedersen, L. L. Clepper, M. Nelson, W. G. Sanger, S. Gokhale, D. P. Wolf & S. M. Mitalipov

In this Article, the legend to Supplementary Fig. 3 should have stated that the image used to illustrate the appearance of a monkey oocyte under polarized microscopy prior to any manipulation is the same illustration as the one originally included in figure 4 of ref. 1. This image is presenting a protocol and not an experimental result.

1. Mitalipov, S. M. Reprogramming following somatic cell nuclear transfer in primates is dependent upon nuclear remodeling. *Hum. Reprod.* **22,** 2232–2242 (2007).

CORRECTIONS & AMENDMENTS

Corrigendum: Nuclear reprogramming by interphase cytoplasm of two-cell mouse embryos

Eunju Kang, Guangming Wu, Hong Ma, Ying Li,
Rebecca Tippner-Hedges, Masahito Tachibana,
Michelle Sparman, Don P. Wolf, Hans R. Schöler
& Shoukhrat Mitalipov

In the first sentence of the main text of this Letter, the words '...*Brg1*, *Bmi1* (also known as *Smarca4*)...' should have read '...*Brg1* (also known as *Smarca4*), *Bmi1*...'. This has been corrected in the PDF and HTML versions of the article.

CORRECTIONS & AMENDMENTS

Corrigendum: Mitochondrial gene replacement in primate offspring and embryonic stem cells

Masahito Tachibana, Michelle Sparman,
Hathaitip Sritanaudomchai, Hong Ma, Lisa Clepper,
Joy Woodward, Ying Li, Cathy Ramsey, Olena Kolotushkina
& Shoukhrat Mitalipov

In this Article, the legend to Fig. 1b should have stated that the image used to illustrate the appearance of a monkey oocyte under polarized microscopy prior to any manipulation is the same illustration as the one originally included in figure 4 of ref. 1. This image is presenting a protocol and not an experimental result.

1. Mitalipov, S. M. Reprogramming following somatic cell nuclear transfer in primates is dependent upon nuclear remodeling. *Hum. Reprod.* **22,** 2232–2242 (2007).

CAREERS

OCEAN BIOLOGY

Marine dreams

Scientists in a glamour field offer tips — and reality checks — for the next generation of marine biologists.

BY CHRIS WOOLSTON

Whenever George Matsumoto gets a call from an unfamiliar number, he has a good idea of who will be on the other end: a young person who dreams about living on a boat and communing with dolphins, whales and otters. As a marine biologist and education specialist at the Monterey Bay Aquarium Research Institute (MBARI) in Moss Landing, California, Matsumoto is a public face for a branch of science that has been glorified and romanticized through films and television shows. Young people from the coast of England to the plains of Kansas are making plans to study sea creatures, and they want to know how to get into the club. "I get phone calls and e-mails non-stop all year long," Matsumoto says. "They're almost always from high-school students. I just got four different e-mails from the same high school in Florida. I don't know how they find me."

Like many researchers, Matsumoto has devoted much of his career to education and mentorship. But as a marine biologist, he is in a tricky position: he has to turn wide-eyed enthusiasm into a grounded understanding of day-to-day research — which often combines the thrill of staring at numbers on a computer screen with the joy of seasickness — without breaking too many spirits.

It is a challenge shared by other marine biologists around the world, whether they are studying tuna or plankton, coral or seaweed. They do not want to discourage anyone from science. But in a field that is already crowded with PhD graduates looking for meaningful work, they want to make sure that the next generation arrives with the right motives and a realistic understanding of the prospects. Newly independent principal investigators who are being chased down by starry-eyed high schoolers and undergraduates should equip themselves with a broad knowledge of education options, a feel for the job market and a deep pool of empathy. After all, they probably once had a few stars in their eyes themselves.

RIGHT MOTIVES

Matsumoto says that he is always happy to make time for those who reach out to him. About a dozen times a year, he will carve an hour out of his schedule to sit down with students who visit the lab. One of his first pieces of advice to callers and e-mailers is for them to check out the website 'So You Want To Be A Marine Biologist?' (see go.nature.com/gqwbum) created by Milton Love, a fisheries researcher at the University of California, Santa Barbara. The site bluntly advises that anyone who wants to become a marine biologist so as to establish some sort of cosmic new-age connection with dolphins should aim for another line of work. "In our experience," it says, "people who feel this way last about 6.5 minutes in any biology program." The site also discourages anyone who wants to get rich from taking up marine biology. "Five years after getting my PhD, I was making slightly less than a beginning manager at McDonalds," Love writes.

Speaking from his office, Love says that despite the warnings on his highly read site, he continues to receive a steady stream of queries from high-school students, undergrads and even people with PhDs in other fields who want to break into ocean science. "I sympathize with these people," he says. "I believe that there's a place in science for anyone with a seeking mind. But I don't want them to get crushed down the road." (Love takes another, more in-depth look at the ins-and-outs of the profession in a follow-up website 'So You Want To Be A Marine Biologist? The Revenge!' (see go.nature.com/utmbiw). ▶

KIMBERLY SWAN

KIMBERLY SWAN

The Watsonville Area Teens for Coastal Habitats programme sets high-school students to work sampling fish in lagoons and estuaries.

Some of the most enthusiastic marine-biologists-to-be have yet to start university. Every year, a group of high-school students visits the University of New Hampshire's Shoals Marine Laboratory on Appledore Island, giving executive director Jennifer Seavey a chance to work with people at the very beginning of the marine-biology pipeline. "It's a field that attracts a lot of young students," she says. "In the 1970s, everyone wanted to become a marine biologist because of Jacques Cousteau." Now, she says, the big draw is *Dolphin Tale*, a 2011 film about a dolphin that receives a prosthetic tail, and Shark Week, a much-hyped binge of shark programmes on the Discovery Channel in the United States.

"The most common thing I hear is that they want to be marine-mammal veterinarians. I tell them that there are maybe five really successful marine-mammal vets in the world," she says. "The rest are techs at SeaWorld", a chain of theme parks in the United States.

Once at the facility, students quickly learn that marine biology does not always follow the heart-warming Hollywood script. Among other endeavours, students get a chance to practice wildlife forensics — taking a close look at dead seals and sea birds, for instance, and trying to work out how they met their demise.

Matsumoto takes groups of students from underserved high schools to field sites along Monterey Bay through the Watsonville Area Teens for Coastal Habitats programme. Almost all the students are Hispanic, and many are still learning English. Language barriers aside, the science is solid. "They pick their own topics," Matsumoto says. "We give them a research site and time to explore, and they come up with their own hypotheses." Ongoing projects include measuring crab density and biodiversity, and identifying plankton. The kids really get into the work, he says, even if it does not exactly fit into their preconceived ideas of ocean research.

The fascination with marine biology is not restricted to high-school students. Many undergraduate students remain enthralled, which explains the pile of applications that MBARI receives for its ten-week internships for university students. "We get 200–300 applicants every year for 12–20 positions," Matsumoto says. Those lucky enough to get an internship are rewarded with a valuable dose of reality. Matsumoto says that they will often have a glorious day of research that is seemingly pulled from the pages of *National Geographic* magazine, then spend weeks and weeks working on the data. "Some of the interns realize it's not for them," he says. "For us, that's a success story." Although for better or for worse, the summer of 2014 had no such 'successes'. "We had humpback whales feeding 200 feet off the beach pretty much all summer," he says. "The interns could watch them during their lunch breaks. After that, none of them wanted to get out of science."

SHARKS AND SEAWEEDS

Andrew Davies, a marine ecologist at Bangor University, UK, is not surprised that so many people want to study the ocean. "It holds incredible biological diversity from the tiniest microbes to the largest organisms on the planet," he says. "And it's not just kids. We have mature students who want to change careers." Whatever their age, the newcomers that he runs across tend to have highly idealized and simplistic ideas of the profession. "The media has developed a myth that now surrounds marine biology, and indeed many careers in the natural sciences," he says. "Students arrive at university with an almost single-minded focus on coral reefs, marine mammals or large predators such as sharks."

One of Davies's jobs, he says, is to show them other possibilities. "I want to expose them to organisms that they've never come across before, such as worms that build large reef-like structures out of sand particles, or long-lived forests of algae that create their own ecosystems." Davies himself started out studying seaweed — a practical choice, he says. "There are far more job opportunities out there on seaweeds than on sharks, often with less competition." But it still took him months to find a job after getting his PhD. "I spent that time working on publications and doing some volunteer work. Now I'm an academic, and I've never looked back. I have loved pretty much every day of my entire career. I work long hours mixing research with teaching, but every day is different."

Competition is a common theme throughout the natural sciences, where the supply of PhD students and postdocs far outstrips the positions in academia. And because so many people want to become marine biologists, university scientists often have to act like gatekeepers. "We can be leery about bringing on graduate students who have their sights set on academia," says Rita Mehta, an evolutionary biologist at the University of California, Santa Cruz's Long Marine Laboratory. "We have to ask ourselves, is this person really ready to fight for a job?"

"Because so many people want to get into the field, you need dedication and creativity."

When talking to undergraduates, she says, she sometimes steers them away from marine biology altogether towards a more general and potentially more marketable degree perhaps in evolution or molecular biology. She says that even at her own institution, ocean science gets an outsized share of student interest even though plenty of terrestrial biologists are doing excellent work. "Marine biology is thought to be the pinnacle of majors, but that's because people don't understand what else is out there."

EDUCATION
Those who can, teach

With so many young people eager to learn about ocean life, marine education can be a promising career path. Whether as a full-time job at an aquarium or at a summer camp on the high seas, explaining marine science to kids can be very rewarding, says Cause Hanna, research manager of the Santa Rosa Island Research Station, part of the California State University Channel Islands. "As a researcher, you can be plugging away on a problem for years," he says. "As an educator, you can get phenomenal results in a day."

According to Jennifer Seavey, executive director of the University of New Hampshire's Shoals Marine Laboratory on Appledore Island, "there are a lot of marine-science camps and courses for kids, and they all need people to teach them". Many of the jobs are at the sorts of places that attract so many people to marine biology. SeaTrek BVI, a company that offers adventure summer camps for teens in the British Virgin Islands, hires biologists to teach kids about coral reefs, mangroves, plankton and other ocean topics. The Marine Discovery Center at New Smyrna Beach in Florida employs biologists to guide dolphin tours, give talks about sharks and starfish to the general public and teach at summer camps for kids and teens.

California's Catalina Island Marine Institute — a non-profit school for children aged 9–17 — is one of the best destinations for early-career marine biologists who have a penchant for teaching, says George Matsumoto, education specialist at the Monterey Bay Aquarium Research Institute in Moss Landing, California. "It has a large network of alumni all over the world," he says. "Having that on your CV will only help you."

For those who prefer more stable work, Seavey notes that a bachelor's or master's degree in ocean science can be a good foundation for a career teaching at pre-university levels. "It's not uncommon to find high-school teachers with a background in marine biology," she says.

Researchers do not necessarily need formal training to share their knowledge with others, but Matsumoto says that it is important to hone teaching skills when you have the chance. "Postdocs should look around at local community colleges to see if they can get an adjunct or guest lecturer position," he says. "PhD students should ask their professors if they can teach some classes. I did that with my professor, and he was more than happy to oblige." **C.W.**

Mehta assures students who are willing to look beyond academia that jobs are out there. "There are quite a few public research opportunities," she says, including positions with aquariums, non-profit organizations and governments at the federal, state and municipal level. Tetra Tech, a consulting firm based in Research Triangle Park, North Carolina, is seeking an aquatic ecologist, and the Alaska Department of Fish and Game in Dutch Harbor wants a fishery biologist, for example. The inexhaustible pool of interest in ocean science among the general public also opens up opportunities for researchers with a penchant for teaching, Mehta adds (see 'Those who can, teach'). If an early-career scientist knows a few things about sea lions, great white sharks or oysters, there will always be people who want to hear about it.

But none of those jobs are easily won. "There are numerous career options," says Erich Hoyt, a researcher with the global non-profit organization Whale and Dolphin Conservation in Chippenham, UK. "But because so many people want to get into the field, you need dedication and creativity." He says that he received more than 200 applications when he recently put out a call for an assistant.

Studying marine mammals in the field requires an especially diverse skill set, Hoyt says. Among other things, he says, researchers need to be able to handle boats of all sizes, take photos, make sound recordings, sort through streams of data and write papers. Hoyt does all these, as well as giving regular talks and writing popular books, including the 2013 children's book *Weird Sea Creatures*, a side career that has undoubtedly sent more young people down a path towards a career in ocean science.

What opportunities will those students have? It depends on the student. "There are no guaranteed jobs post-graduation in any field, especially in a competitive area such as marine biology," Davies says. But the picture is not hopeless. "There is always a need for enthusiastic, motivated and hard-working graduates who have the confidence to tackle challenges head on." If that challenge involves spotting blue whales from a boat or scuba diving with a pod of dolphins, so be it. It is a tough job, but some marine biologist will have to do it. ∎

Chris Woolston *is a landlocked freelance writer in Billings, Montana.*

Gloomy outlook

US universities will probably face financial pressure until at least mid-2016, including an erosion of federal funding, says a report by Moody's Investors Service in New York. The report, *2015 Outlook — US Higher Education: Slow Tuition Revenue Growth Supports Negative Outlook*, released on 1 December, predicts that universities will continue to battle for tuition-fee revenue, state funding and federal grants. Moody's, a credit-rating agency, expects federal grant amounts and activity, especially from the US National Institutes of Health and the US National Science Foundation, to decline in the next 12–18 months. It says that the contraction will be a result of discretionary spending cuts, federal budget pressures and the continuing effects of last year's across-the-board funding sequestration. Research will increasingly be funded through private donations and gifts, the report predicts. The continued negative outlook, in effect since January 2013, means that Moody's is more likely to give poor credit ratings to US universities, which will incur higher borrowing costs and might be forced to scale back hiring plans.

Informal relations

Women are more likely to realize career benefits from informal relationships with colleagues and others if they are in a discipline that comprises at least 15% women and are not simply tokens, finds a study. *Cultural Correlates of Gender Integration in Science* analysed accounts of scientific success in psychology, psychiatry and the life sciences, which have large proportions of women, and in engineering and physics, in which women tend to be underrepresented. The authors found that informal relationships (including those with colleagues and contacts made through conferences or other means) help women to integrate and stay in their career just as much as mentorships and other formally structured relationships. They suggest that the benefits come from the extra support and opportunities these relationships can provide. Early-career female researchers should assess the collegiality of their fields and workplaces as they make career decisions, says co-author Cindy Cain, a postdoc at the University of Minnesota in Minneapolis. "Friendly relationships may increase women's sense of professional role confidence, thus helping them to fit in and be productive — as long as women have surpassed the 15% tokenism level in that discipline," Cain says.

EDUCATION
Those who can, teach

With so many young people eager to learn about ocean life, marine education can be a promising career path. Whether as a full-time job at an aquarium or at a summer camp on the high seas, explaining marine science to kids can be very rewarding, says Cause Hanna, research manager of the Santa Rosa Island Research Station, part of the California State University Channel Islands. "As a researcher, you can be plugging away on a problem for years," he says. "As an educator, you can get phenomenal results in a day."

According to Jennifer Seavey, executive director of the University of New Hampshire's Shoals Marine Laboratory on Appledore Island, "there are a lot of marine-science camps and courses for kids, and they all need people to teach them". Many of the jobs are at the sorts of places that attract so many people to marine biology. SeaTrek BVI, a company that offers adventure summer camps for teens in the British Virgin Islands, hires biologists to teach kids about coral reefs, mangroves, plankton and other ocean topics. The Marine Discovery Center at New Smyrna Beach in Florida employs biologists to guide dolphin tours, give talks about sharks and starfish to the general public and teach at summer camps for kids and teens.

California's Catalina Island Marine Institute — a non-profit school for children aged 9–17 — is one of the best destinations for early-career marine biologists who have a penchant for teaching, says George Matsumoto, education specialist at the Monterey Bay Aquarium Research Institute in Moss Landing, California. "It has a large network of alumni all over the world," he says. "Having that on your CV will only help you."

For those who prefer more stable work, Seavey notes that a bachelor's or master's degree in ocean science can be a good foundation for a career teaching at pre-university levels. "It's not uncommon to find high-school teachers with a background in marine biology," she says.

Researchers do not necessarily need formal training to share their knowledge with others, but Matsumoto says that it is important to hone teaching skills when you have the chance. "Postdocs should look around at local community colleges to see if they can get an adjunct or guest lecturer position," he says. "PhD students should ask their professors if they can teach some classes. I did that with my professor, and he was more than happy to oblige." **C.W.**

Mehta assures students who are willing to look beyond academia that jobs are out there. "There are quite a few public research opportunities," she says, including positions with aquariums, non-profit organizations and governments at the federal, state and municipal level. Tetra Tech, a consulting firm based in Research Triangle Park, North Carolina, is seeking an aquatic ecologist, and the Alaska Department of Fish and Game in Dutch Harbor wants a fishery biologist, for example. The inexhaustible pool of interest in ocean science among the general public also opens up opportunities for researchers with a penchant for teaching, Mehta adds (see 'Those who can, teach'). If an early-career scientist knows a few things about sea lions, great white sharks or oysters, there will always be people who want to hear about it.

But none of those jobs are easily won. "There are numerous career options," says Erich Hoyt, a researcher with the global non-profit organization Whale and Dolphin Conservation in Chippenham, UK. "But because so many people want to get into the field, you need dedication and creativity." He says that he received more than 200 applications when he recently put out a call for an assistant.

Studying marine mammals in the field requires an especially diverse skill set, Hoyt says. Among other things, he says, researchers need to be able to handle boats of all sizes, take photos, make sound recordings, sort through streams of data and write papers. Hoyt does all these, as well as giving regular talks and writing popular books, including the 2013 children's book *Weird Sea Creatures*, a side career that has undoubtedly sent more young people down a path towards a career in ocean science.

What opportunities will those students have? It depends on the student. "There are no guaranteed jobs post-graduation in any field, especially in a competitive area such as marine biology," Davies says. But the picture is not hopeless. "There is always a need for enthusiastic, motivated and hard-working graduates who have the confidence to tackle challenges head on." If that challenge involves spotting blue whales from a boat or scuba diving with a pod of dolphins, so be it. It is a tough job, but some marine biologist will have to do it. ∎

Chris Woolston *is a landlocked freelance writer in Billings, Montana.*

Gloomy outlook

US universities will probably face financial pressure until at least mid-2016, including an erosion of federal funding, says a report by Moody's Investors Service in New York. The report, *2015 Outlook — US Higher Education: Slow Tuition Revenue Growth Supports Negative Outlook*, released on 1 December, predicts that universities will continue to battle for tuition-fee revenue, state funding and federal grants. Moody's, a credit-rating agency, expects federal grant amounts and activity, especially from the US National Institutes of Health and the US National Science Foundation, to decline in the next 12–18 months. It says that the contraction will be a result of discretionary spending cuts, federal budget pressures and the continuing effects of last year's across-the-board funding sequestration. Research will increasingly be funded through private donations and gifts, the report predicts. The continued negative outlook, in effect since January 2013, means that Moody's is more likely to give poor credit ratings to US universities, which will incur higher borrowing costs and might be forced to scale back hiring plans.

CAREER PROGRESS
Informal relations

Women are more likely to realize career benefits from informal relationships with colleagues and others if they are in a discipline that comprises at least 15% women and are not simply tokens, finds a study. *Cultural Correlates of Gender Integration in Science* analysed accounts of scientific success in psychology, psychiatry and the life sciences, which have large proportions of women, and in engineering and physics, in which women tend to be underrepresented. The authors found that informal relationships (including those with colleagues and contacts made through conferences or other means) help women to integrate and stay in their career just as much as mentorships and other formally structured relationships. They suggest that the benefits come from the extra support and opportunities these relationships can provide. Early-career female researchers should assess the collegiality of their fields and workplaces as they make career decisions, says co-author Cindy Cain, a postdoc at the University of Minnesota in Minneapolis. "Friendly relationships may increase women's sense of professional role confidence, thus helping them to fit in and be productive — as long as women have surpassed the 15% tokenism level in that discipline," Cain says.

MISSED MESSAGE

And finally ...

BY RACHEL REDDICK

This wasn't what I had planned for today. I had planned to watch the night sky. Instead, I'm waiting for the end of the world. The people on the high-orbit station tell me that I've got less than half an hour before the fireball gets here.

The low-orbit station may have been hit by debris. The high station says they're not responding to signals, so we're assuming the worst.

It's annoying. If they'd have survived, maybe we'd have a better idea of what was about to kill us all. Someone on the far side of the planet must have done something wrong. There had been rumours. One country claimed that they were constructing a faster-than-light interstellar drive, which most people assumed was a cover story for some kind of super-bomb. All that tritium, deuterium and plutonium they were collecting had to be for something. But this … who would build this? Based on the size, one of the astronauts estimated that the explosion was nearly big enough to disrupt the entire planet. Were they playing with antimatter and lost containment? They'd have needed an immense amount of antimatter to do this. How could they get anything more than a handful of particles?

I guess we'll never know. It's an impossible disaster. An explosion like this … it'll scour the surface clean, just as if a huge asteroid had hit. Except that we would have seen an asteroid that big coming.

No, not an asteroid. It'd need to be practically a dwarf planet. And there aren't any of those close enough to do anything to us.

As always, it's not nature that hurts us most. We're our own most effective exterminators.

I don't want to die.

An interstellar drive would be really useful right now.

But I don't have one, or even a basic launch vehicle. So I'm making the best with what I've got. I'm lucky enough, relatively speaking, to be far from the epicentre of the disaster, sitting in the control room of one of our best radio telescopes. We use … *used* it to track potentially dangerous asteroids, sending out a strong pulse of radio waves and listening for the echo. We watched the skies for threats that could destroy cities. Maybe we should have been looking down, instead of up.
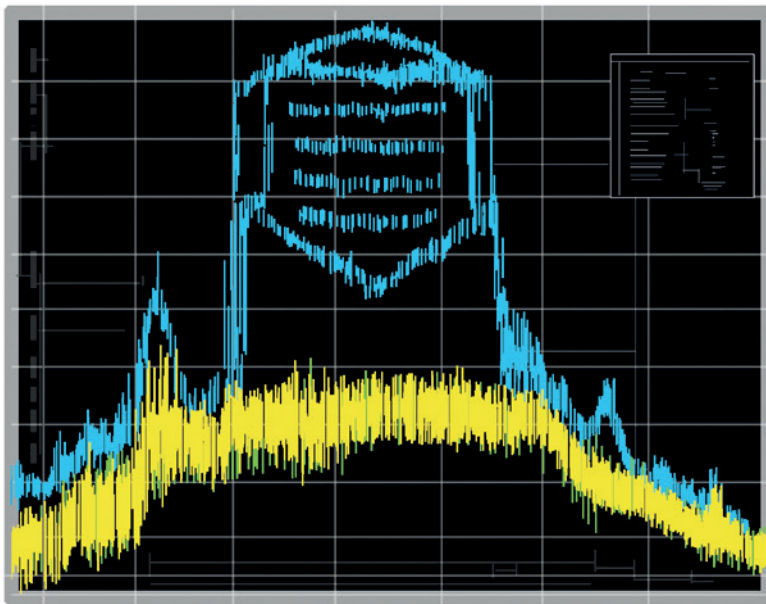
Either way, this means I have one of the biggest and best radio transmitters at my fingertips. I may die, all of the people I have ever known and loved may die, but I can at least ensure that we will not be forgotten. That's why I'm sending this message.

Not that anyone will know how to translate it. And I suppose that there's a good chance nobody will even hear it. The telescope's radar beam is fairly narrow, as it was designed for locating asteroids. I'd aim for the neighbouring galaxy, if I could. The beam is wide enough at that distance to hit a lot of stars, and still be strong enough for someone to notice. But I can't send a radio signal through the ground.

Even if I do manage to reach an inhabited system, there's a chance they won't even be looking when the message comes by. Maybe just sending the message is enough. Enough to know that the Universe will always bear our mark, travelling forever through space. It's better than doing nothing.

I don't know what the astronauts are going to do. Most of their homes have already been vaporized. In a few minutes, there isn't going to be anything on the surface colder than molten lava. They're going to have to fire the thrusters, to avoid the debris that will soon be in orbit. Even if they do, the station isn't a closed system. They're going to run out of supplies eventually. Maybe they'll wait until they slowly starve to death. Maybe they'll decide that moving the station isn't worth the effort.

I'm glad I don't have that decision to make. I'm trying to work out if the rumbling I feel is the first sign of the approaching shockwave, or just me shaking from my own fear. I suppose I'll be signing off, whether I want to or not.

Until then, I'll keep transmitting. A last call, before we are forever silenced.

Several hundred years later…

"Huh." The young woman spun in her chair. "Jordan, can you have a look at this?"

"Have a look at what?" He looked at where her finger traced a series of lines on the screen, all drifting up and down chaotically.

"Do you see a signal there?"

"No, not a thing. Why?"

"Remember the unusually strong blip near the beginning of observations a couple of days ago? It wasn't far from the hydrogen lines. The short track I got that moved with Earth's rotation?"

He nodded. "Which is why it got follow-up. But the original signal didn't last very long."

"A few minutes. And when I looked there tonight, and last night, there's just nothing. Nothing at all."

Jordan gave her a sympathetic look. "That's too bad. You think it was a message?"

She tapped at the keyboard, pulling up a slice of the earlier observations. The spike was obvious. "It was narrow, it was near the hydrogen line … but it didn't last long enough to confirm one way or another," she said sadly. "We'll probably never know." ■

Rachel Reddick *recently completed her graduate work in physics, and she currently teaches at Foothill Community College. She enjoys fiction and storytelling in her spare time.*
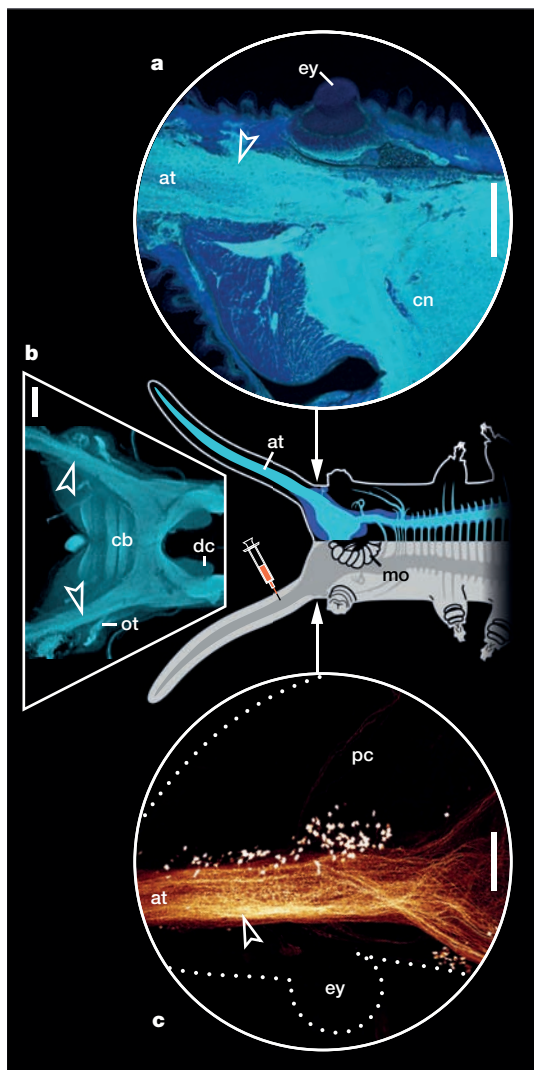
ILLUSTRATION BY JACEY
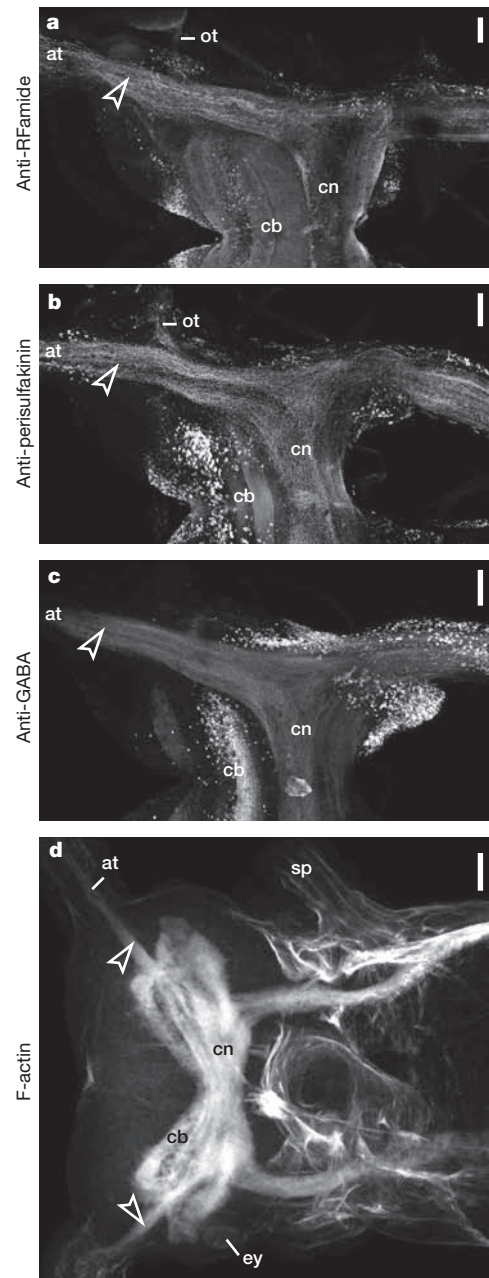
Latest anomalocaridid affinities challenged

Cong *et al.*[1] report a new anomalocaridid species, *Lyrarapax unguispinus,* that bears a potential pair of pre-protocerebral ganglia associated with frontal appendages, thus challenging some previous assignments of these appendages to the second (deutocerebral) segment[2,3]. On the basis of putative similarities in brain anatomy to the extant onychophoran *Euperipatoides rowelli*, the authors go further by assigning homology between the anomalocaridid-like appendages and the arthropod labrum[4]. However, we demonstrate that their arguments are based on a misinterpretation of onychophoran neuroanatomy. Consequently, we believe that the proposed affinities of these appendages are incorrect

and their homologues remain uncertain. There is a Reply to this Brief Communication Arising by Cong, P. *et al. Nature* **516,** http://dx.doi.org/10.1038/nature13861 (2014).

Figure 1 | Head and brain anatomy in *Euperipatoides rowelli,* the same species studied by Cong et al.[1]. a–c, Confocal micrographs and diagram. The brain lies dorsal rather than anterior to the mouth. Arrowheads indicate the expected position of the putative 'pre-protocerebral ganglia'. **a,** Vibratome section of the dorsal brain labelled with a DNA marker (dark blue) and acetylated α-tubulin (cyan). **b,** Maximum projection of the brain in dorsal view (anti-synapsin immunolabelling). **c,** Antennal tract filled with fluorescein-tagged dextran (syringe indicates fill site). at, antennal tract; cb, central body; cn, central brain neuropil; dc, deutocerebrum; ey, eye; mo, mouth; ot, optic tract; pc, protocerebrum. Scale bars: **a, b,** 200 μm; **c,** 100 μm.

Figure 2 | Onychophoran brain anatomy. Confocal micrographs of adult brains and an embryonic head labelled with different markers in dorsal view. Anterior is to the left. Note that no pre-protocerebral ganglia are present in the position described by Cong *et al.*[1] (arrowheads). **a,** Adult brain of the peripatid *Principapillatus hitoyensis.* **b, c,** Adult brains of the peripatopsid *Euperipatoides rowelli.* **d,** Embryonic head of *E. rowelli* (late stage V embryo). at, antennal tract; cb, central body; cn, central brain neuropil; ey, developing eye; ot, optic tract; sp, slime papilla. Scale bars: **a–c,** 100 μm; **d,** 50 μm.

BRIEF COMMUNICATIONS ARISING

First, regardless of whether or not remnants of frontal appendage ganglia were present in *L. unguispinus*, developmental and neuroanatomical data[5–9] clearly show that 'pre-protocerebral ganglia' do not exist in onychophorans (Figs 1 and 2). We believe that the structure labelled 'frontal appendage ganglion' by Cong *et al.*[1] in brain sections of *E. rowelli* is a portion of the antennal tract (arrowheads in Figs 1a–c and 2b–d) situated in the anterior protocerebrum[5,8]. Unfortunately, the single section shown in Fig. 3a of ref. 1 is misleading, as it excludes the antennal tract, making the putative 'frontal appendage ganglion' appear as a separate structure. Furthermore, the regions of the antennal tract labelled 'frontal appendage nerve' and 'frontal ganglion' in Extended Data Fig. 2e of ref. 1 are cytologically indistinct from each other and from the remaining tract in properly sectioned specimens[7] and whole-mount preparations of brains and heads (Figs 1 and 2). We therefore conclude that the slight indentation of the antennal tract seen in Extended Data Fig. 2e of ref. 1 is most likely a sectioning artefact, which is not evident in previous specimens that were prepared using the same technique[7]. Full confocal projections of onychophoran brains and heads using different markers demonstrate no difference between the 'ganglion-like neuropil' region of Cong *et al.*[1] and the remaining portions of the antennal tract (Figs 1 and 2). In our view, the authors' limited data do not permit the unequivocal interpretation of the true shape of neuropils and nerve tracts of the onychophoran brain.

Second, the onychophoran antennae are protocerebral rather than pre-protocerebral appendages because, despite interneurons associated with the medullary antennal tracts, their supplying neurons are clearly concentrated within the brain[8] (Fig. 1c). Even if ganglion-like structures were present in *E. rowelli*, the sections in Cong *et al.*[1] indicate that they would be situated within rather than anterior to the protocerebrum (this is evident from Fig. 3d of ref. 1, in which the light-grey regions seem to represent the true contours of the onychophoran brain). This contradicts not only their pre-protocerebral position but also their designation as ganglia, since a ganglion itself cannot logically encompass another ganglion. Furthermore, the lack of antennal ganglia corresponds to the general absence of limb ganglia in Onychophora[10], as the antennae themselves are modified limbs[11].

Third, the claim that the onychophoran brain consists of a single segment is challenged by neuronal tracing data of the jaw nerves, which show that the onychophoran brain is undoubtedly a bipartite structure[8]. The argument of Cong *et al.*[1] based solely on previous *engrailed* messenger RNA (incorrectly referenced as protein) expression data overlooks the fact that the anterior *engrailed* stripe is on the non-neuroectodermal side adjacent to the invaginating eye, rendering these data irrelevant for addressing brain segmentation[12,13].

Fourth, Cong *et al.*[1] claim to clarify the homology and affinity of anomalocaridid frontal appendages and onychophoran antennae by using the position of the eyes as an anatomical landmark. However, we believe that this comparison is inadequate for resolving the spatial relationship of the frontal appendages as pre-ocular structures because final position does not necessarily indicate segmental origin. This is evident from frontal appendages of arthropods, for example, brine shrimp and house centipedes[14], which are also positioned anterior to the eyes yet are innervated by the deutocerebrum. Therefore, the physical position of the anomalocaridid frontal appendages is inappropriate for deciphering their segmental identity.

On the basis of these conflicting data and previous evidence[5–9], the presented scenario[1] ceases to be viable. The data of Cong *et al.*[1] do not support the homology of the onychophoran antennae with the frontal appendages of *L. unguispinus* and therefore are irrelevant for resolving the segmental affinity of these appendages. Consequently, the 'hypothetical ancestor' used as a basis for resolving the anomalocaridid frontal appendages as homologues of the arthropod labrum is a tenuous speculation.

Georg Mayer[1], Christine Martin[1], Ivo de Sena Oliveira[1], Franziska Anni Franke[1] & Vladimir Gross[1]

[1]Animal Evolution and Development, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany.
email: gmayer@onychophora.com

1. Cong, P., Ma, X., Hou, X., Edgecombe, G. D. & Strausfeld, N. J. Brain structure resolves the segmental affinity of anomalocaridid appendages. *Nature* **513,** 538–542 (2014).
2. Stein, M. A new arthropod from the Early Cambrian of North Greenland, with a 'great appendage'-like antennula. *Zool. J. Linn. Soc.* **158,** 477–500 (2010).
3. Haug, J. T., Waloszek, D., Maas, A., Liu, Y. U. & Haug, C. Functional morphology, ontogeny and evolution of mantis shrimp-like predators in the Cambrian. *Palaeontology* **55,** 369–399 (2012).
4. Budd, G. E. A palaeontological solution to the arthropod head problem. *Nature* **417,** 271–275 (2002).
5. Schürmann, F. W. in *Arthropod Brain, its Evolution, Development, Structure, and Functions* (ed. Gupta, A. P.) Ch. 8, 159–180 (John Wiley & Sons, 1987).
6. Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanangrensis* with particular reference to the central nervous system. *J. Morphol.* **255,** 1–23 (2003).
7. Strausfeld, N. J., Strausfeld, C., Stowe, S., Rowell, D. & Loesel, R. The organization and evolutionary implications of neuropils and their neurons in the brain of the onychophoran *Euperipatoides rowelli*. *Arthropod Struct. Dev.* **35,** 169–196 (2006).
8. Mayer, G., Whitington, P. M., Sunnucks, P. & Pflüger, H.-J. A revision of brain composition in Onychophora (velvet worms) suggests that the tritocerebrum evolved in arthropods. *BMC Evol. Biol.* **10,** 255 (2010).
9. Whitington, P. M. & Mayer, G. The origins of the arthropod nervous system: Insights from the Onychophora. *Arthropod Struct. Dev.* **40,** 193–209 (2011).
10. Mayer, G. & Harzsch, S. Immunolocalization of serotonin in Onychophora argues against segmental ganglia being an ancestral feature of arthropods. *BMC Evol. Biol.* **7,** 118 (2007).
11. Mayer, G. & Koch, M. Ultrastructure and fate of the nephridial anlagen in the antennal segment of *Epiperipatus biolleyi* (Onychophora, Peripatidae)—evidence for the onychophoran antennae being modified legs. *Arthropod Struct. Dev.* **34,** 471–480 (2005).
12. Hirth, F. *et al.* An urbilaterian origin of the tripartite brain: developmental genetic insights from *Drosophila*. *Development* **130,** 2365–2373 (2003).
13. Eriksson, B., Samadi, L. & Schmid, A. The expression pattern of the genes *engrailed, pax6, otd* and *six3* with special respect to head and eye development in *Euperipatoides kanangrensis* Reid 1996 (Onychophora: Peripatopsidae). *Dev. Genes Evol.* **223,** 237–246 (2013).
14. Sombke, A., Harzsch, S. & Hansson, B. S. Organization of deutocerebral neuropils and olfactory behavior in the centipede *Scutigera coleoptrata* (Linnaeus, 1758) (Myriapoda: Chilopoda). *Chem. Senses* **36,** 43–61 (2011).

BRIEF COMMUNICATIONS ARISING

Cong *et al.* reply

In the accompanying Comment Mayer *et al.*[1] dispute ancestral affinities of *Lyrarapax unguispinus* and onychophoran brains. Here, we contest their claim that the evolutionary scenario described in Cong *et al.*[2] is unviable.

Mayer *et al.*[1] suggest that the frontal appendage ganglion in Extended Data Fig. 2e of ref. 2 is a 'sectioning artefact'. To demonstrate the frontal appendage ganglion, we present three consecutive silver-stained sections (Fig. 1a–c) of specimens from the data set[3] that Mayer *et al.*[1] accept as

Figure 1 | *Euperipatoides rowelli* frontal appendage ganglia. Hemi-brains; anterior is up. **a–c,** Three consecutive silver-stained sections (ventral to dorsal) showing frontal appendage nerve bundles (fan) merging with the frontal appendage ganglion (frg) anterior to the eye (ey) and optic tract (opt), which supplies nested optic neuropils (on2) flanking the medial protocerebrum's (mpr) central complex (cc). The frontal appendage ganglion is confluent with the medial protocerebrum (border indicated with an arrow in **b**). Insets in **c** distinguish axons (fan) and neuropil (frg). **d,** Osmium-ethyl gallate-stained brain showing axons from the fan entering the frontal appendage ganglion and a broad axon fascicle (bracket) leaving it to flank the lateral protocerebrum (lpr) before entering the lateral nerve cord (lat nc), which corresponds to one of the paired descending tracts of *L. unguispinus* (dt in Fig. 2g of Cong *et al.*[2]). Bundled neurites (mb) supply the mushroom bodies, which originate ventral to this level. A small volume of mb neuropil is visible in panel **a**. Scale bars, 100 μm.

'properly sectioned'. Figure 1c demonstrates the frontal appendage ganglion as being cytologically distinct from the nerve bundles entering it. Figure 1d shows axons entering it frontally and extending from it caudally. To illustrate that the frontal appendage ganglia do not exist, Mayer *et al.*[1] offer their Fig. 1a, which although labelled with anti-α-tubulin lacks the resolution to distinguish neuropil from axons. Their Figs 1b and 2a–d, claimed as 'full confocal projections', omit from the brain dorsal neuropils, optic tracts with optic neuropils, mushroom bodies (the brain's most prominent synaptic neuropil[3]), and lateral protocerebral regions. Their confocal images demonstrate that antisera useful for revealing axonal tracts, or selectively resolving peptidergic systems, fail to show neuropils revealed by silver or osmium-ethyl gallate staining[2,3].

To support their claim[1] that no neurons exist anterior to the protocerebrum, they offer Fig. 1c, in which sparse neuronal perikarya are stained presumably by dye leakage. Frontal appendages ('antennae' in Fig. 2b of ref. 1) show perikarya in front of the label 'at'. Developmental studies[4,5] also describe neurons and neuropil extending into the frontal appendage base.

Mayer *et al.*[1] state that if frontal appendage ganglia did exist they would be situated within the protocerebrum. Their statement that 'antennae' are protocerebral conflicts with a previous report[6] assigning to them a separate segment, thereby setting a precedent for introducing the term 'pre-protocerebral'. Mayer *et al.*[1] add confusion by stating that appendages, including the 'antennae', do not relate to segmental ganglia, citing the onychophoran central nervous system as asegmental[7]. If they subsequently claim[8] the protocerebrum as a brain segment, then appendages supplying it relate to that segment. The same applies to the onychophoran jaws, the nerves of which arise from the deutocerebrum[8] and whose claw-like shape[9] reveals an appendicular development and ancestry[10,11]. Identically formed claws define walking legs of Cambrian stem-group onychophorans[9]. However, onychophoran frontal appendages conspicuously lack claws, along with certain genes expressed in other appendages[12], suggesting that frontal appendages are either derived or that they antecede the evolution of other appendages.

Citing a pre-ocular location of chilopod antennae[13], Mayer *et al.*[1] argue that eye position is inadequate for resolving segmental identities. However, segmental affiliation relates not to an appendage's final post-developmental location on the head but to the brain segment supplying its axons. Chilopod antennae are supplied from the deutocerebrum. In Onychophora, frontal appendage axons supply paired centres anterior to the optic nerves, which define the protocerebrum. Hence frontal appendage centres are pre-protocerebral and accord with a frontal 'antennal' segment[6].

Peiyun Cong[1], **Xiaoya Ma**[1,2], **Xianguang Hou**[1], **Gregory D. Edgecombe**[2] **& Nicholas J. Strausfeld**[3,4]

[1]Yunnan Key Laboratory for Palaeobiology, Yunnan University, Kunming 650091, China.
email: xghou@ynu.edu.cn
[2]Department of Earth Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
[3]Department of Neuroscience, University of Arizona, Tucson, Arizona 85721, USA.
[4]Center for Insect Science, University of Arizona, Tucson, Arizona 85721, USA.
email: flybrain@neurobio.arizona.edu

1. Mayer, G., Martin, C., Oliveira, I. S., Franke, F. A. & Gross, V. Latest anomalocaridid affinities challenged. *Nature* **516**, http://dx.doi.org/10.1038/nature13860 (2014).

2. Cong, P., Ma, X., Hou, X., Edgecombe, G. D. & Strausfeld, N. J. Brain structure resolves the segmental affinity of anomalocaridid appendages. *Nature* **513,** 538–542 (2014).

3. Strausfeld, N. J., Strausfeld, C. M., Stowe, S., Rowell, D. & Loesel, R. The organization and evolutionary implications of neuropils and their neurons in the brain of the onychophoran *Euperipatoides rowelli. Arthropod Struct. Dev.* **35,** 169–196 (2006).

4. Eriksson, B. J. & Budd, G. E. Onychophoran cephalic nerves and their bearing on our understanding of head segmentation and stem-group evolution of Arthropoda. *Arthropod Struct. Dev.* **29,** 197–209 (2000).

5. Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanangrensis* with particular reference to the central nervous system. *J. Morphol.* **255,** 1–23 (2003).

6. Treffkorn, S. & Mayer, G. Expression of the decapentaplegic ortholog in embryos of the onychophoran *Euperipatoides rowelli. Gene Expr. Patterns* **13,** 384–394 (2013).

7. Mayer, G. & Harzsch, S. Immunolocalization of serotonin in Onychophora argues against segmental ganglia being an ancestral feature of arthropods. *BMC Evol. Biol.* **7,** 118 (2007).

8. Mayer, G., Whitington, P. M., Sunnucks, P. & Pfluger, H. J. A revision of brain composition in Onychophora (velvet worms) suggests that the tritocerebrum evolved in arthropods. *BMC Evol. Biol.* **10,** 255 (2010).

9. Smith, M. R. & Ortega-Hernández, J. *Hallucigenia*'s onychophoran-like claws and the case for Tactopoda. *Nature* **514,** 363–366 (2014).

10. Kennel von, J. Entwicklungsgeschichte von *Peripatus edwardsii* Blanch. und *Peripatus torquatus* n. sp. *Arch. Zool.-Anat. Inst. Würzburg* **8,** 1e93 (1888).

11. Sedgwick, A. *Peripatus* (Macmillan, 1922).

12. Oliveira, M. B. *et al.* Expression of arthropod distal limb-patterning genes in the onychophoran *Euperipatoides kanangrensis. Dev. Genes Evol.* **224,** 87–96 (2014).

13. Sombke, A., Harzsch, S. & Hansson, B. S. Organization of deutocerebral neuropils and olfactory behavior in the centipede *Scutigera coleoptrata* (Linnaeus, 1758) (Myriapoda: Chilopoda). *Chem. Senses* **36,** 43–61 (2011).
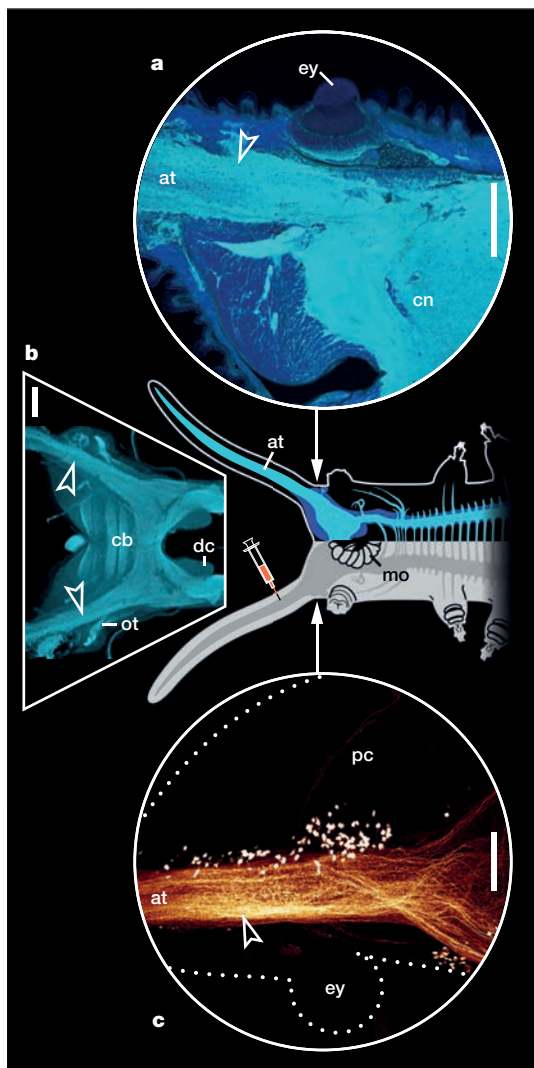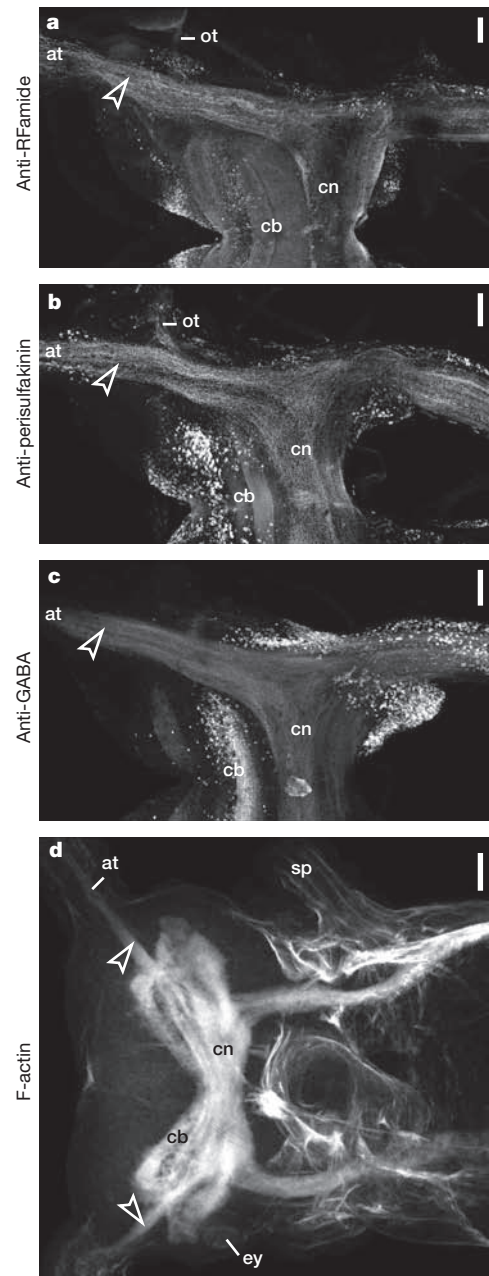
Latest anomalocaridid affinities challenged

Cong *et al.*[1] report a new anomalocaridid species, *Lyrarapax unguispinus,* that bears a potential pair of pre-protocerebral ganglia associated with frontal appendages, thus challenging some previous assignments of these appendages to the second (deutocerebral) segment[2,3]. On the basis of putative similarities in brain anatomy to the extant onychophoran *Euperipatoides rowelli*, the authors go further by assigning homology between the anomalocaridid-like appendages and the arthropod labrum[4]. However, we demonstrate that their arguments are based on a misinterpretation of onychophoran neuroanatomy. Consequently, we believe that the proposed affinities of these appendages are incorrect

and their homologues remain uncertain. There is a Reply to this Brief Communication Arising by Cong, P. *et al. Nature* **516,** http://dx.doi.org/10.1038/nature13861 (2014).

Figure 1 | Head and brain anatomy in *Euperipatoides rowelli,* the same species studied by Cong *et al.*[1]. a–c, Confocal micrographs and diagram. The brain lies dorsal rather than anterior to the mouth. Arrowheads indicate the expected position of the putative 'pre-protocerebral ganglia'. **a,** Vibratome section of the dorsal brain labelled with a DNA marker (dark blue) and acetylated α-tubulin (cyan). **b,** Maximum projection of the brain in dorsal view (anti-synapsin immunolabelling). **c,** Antennal tract filled with fluorescein-tagged dextran (syringe indicates fill site). at, antennal tract; cb, central body; cn, central brain neuropil; dc, deutocerebrum; ey, eye; mo, mouth; ot, optic tract; pc, protocerebrum. Scale bars: **a, b,** 200 μm; **c,** 100 μm.

Figure 2 | Onychophoran brain anatomy. Confocal micrographs of adult brains and an embryonic head labelled with different markers in dorsal view. Anterior is to the left. Note that no pre-protocerebral ganglia are present in the position described by Cong *et al.*[1] (arrowheads). **a,** Adult brain of the peripatid *Principapillatus hitoyensis*. **b, c,** Adult brains of the peripatopsid *Euperipatoides rowelli*. **d,** Embryonic head of *E. rowelli* (late stage V embryo). at, antennal tract; cb, central body; cn, central brain neuropil; ey, developing eye; ot, optic tract; sp, slime papilla. Scale bars: **a–c,** 100 μm; **d,** 50 μm.

BRIEF COMMUNICATIONS ARISING

First, regardless of whether or not remnants of frontal appendage ganglia were present in *L. unguispinus*, developmental and neuroanatomical data[5–9] clearly show that 'pre-protocerebral ganglia' do not exist in onychophorans (Figs 1 and 2). We believe that the structure labelled 'frontal appendage ganglion' by Cong *et al.*[1] in brain sections of *E. rowelli* is a portion of the antennal tract (arrowheads in Figs 1a–c and 2b–d) situated in the anterior protocerebrum[5,8]. Unfortunately, the single section shown in Fig. 3a of ref. 1 is misleading, as it excludes the antennal tract, making the putative 'frontal appendage ganglion' appear as a separate structure. Furthermore, the regions of the antennal tract labelled 'frontal appendage nerve' and 'frontal ganglion' in Extended Data Fig. 2e of ref. 1 are cytologically indistinct from each other and from the remaining tract in properly sectioned specimens[7] and whole-mount preparations of brains and heads (Figs 1 and 2). We therefore conclude that the slight indentation of the antennal tract seen in Extended Data Fig. 2e of ref. 1 is most likely a sectioning artefact, which is not evident in previous specimens that were prepared using the same technique[7]. Full confocal projections of onychophoran brains and heads using different markers demonstrate no difference between the 'ganglion-like neuropil' region of Cong *et al.*[1] and the remaining portions of the antennal tract (Figs 1 and 2). In our view, the authors' limited data do not permit the unequivocal interpretation of the true shape of neuropils and nerve tracts of the onychophoran brain.

Second, the onychophoran antennae are protocerebral rather than pre-protocerebral appendages because, despite interneurons associated with the medullary antennal tracts, their supplying neurons are clearly concentrated within the brain[8] (Fig. 1c). Even if ganglion-like structures were present in *E. rowelli*, the sections in Cong *et al.*[1] indicate that they would be situated within rather than anterior to the protocerebrum (this is evident from Fig. 3d of ref. 1, in which the light-grey regions seem to represent the true contours of the onychophoran brain). This contradicts not only their pre-protocerebral position but also their designation as ganglia, since a ganglion itself cannot logically encompass another ganglion. Furthermore, the lack of antennal ganglia corresponds to the general absence of limb ganglia in Onychophora[10], as the antennae themselves are modified limbs[11].

Third, the claim that the onychophoran brain consists of a single segment is challenged by neuronal tracing data of the jaw nerves, which show that the onychophoran brain is undoubtedly a bipartite structure[8]. The argument of Cong *et al.*[1] based solely on previous *engrailed* messenger RNA (incorrectly referenced as protein) expression data overlooks the fact that the anterior *engrailed* stripe is on the non-neuroectodermal side adjacent to the invaginating eye, rendering these data irrelevant for addressing brain segmentation[12,13].

Fourth, Cong *et al.*[1] claim to clarify the homology and affinity of anomalocaridid frontal appendages and onychophoran antennae by using the position of the eyes as an anatomical landmark. However, we believe that this comparison is inadequate for resolving the spatial relationship of the frontal appendages as pre-ocular structures because final position does not necessarily indicate segmental origin. This is evident from frontal appendages of arthropods, for example, brine shrimp and house centipedes[14], which are also positioned anterior to the eyes yet are innervated by the deutocerebrum. Therefore, the physical position

of the anomalocaridid frontal appendages is inappropriate for deciphering their segmental identity.

On the basis of these conflicting data and previous evidence[5–9], the presented scenario[1] ceases to be viable. The data of Cong *et al.*[1] do not support the homology of the onychophoran antennae with the frontal appendages of *L. unguispinus* and therefore are irrelevant for resolving the segmental affinity of these appendages. Consequently, the 'hypothetical ancestor' used as a basis for resolving the anomalocaridid frontal appendages as homologues of the arthropod labrum is a tenuous speculation.

Georg Mayer[1], Christine Martin[1], Ivo de Sena Oliveira[1], Franziska Anni Franke[1] & Vladimir Gross[1]
[1]Animal Evolution and Development, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany.
email: gmayer@onychophora.com

1. Cong, P., Ma, X., Hou, X., Edgecombe, G. D. & Strausfeld, N. J. Brain structure resolves the segmental affinity of anomalocaridid appendages. *Nature* **513**, 538–542 (2014).
2. Stein, M. A new arthropod from the Early Cambrian of North Greenland, with a 'great appendage'-like antennula. *Zool. J. Linn. Soc.* **158**, 477–500 (2010).
3. Haug, J. T., Waloszek, D., Maas, A., Liu, Y. U. & Haug, C. Functional morphology, ontogeny and evolution of mantis shrimp-like predators in the Cambrian. *Palaeontology* **55**, 369–399 (2012).
4. Budd, G. E. A palaeontological solution to the arthropod head problem. *Nature* **417**, 271–275 (2002).
5. Schürmann, F. W. in *Arthropod Brain, its Evolution, Development, Structure, and Functions* (ed. Gupta, A. P.) Ch. 8, 159–180 (John Wiley & Sons, 1987).
6. Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanangrensis* with particular reference to the central nervous system. *J. Morphol.* **255**, 1–23 (2003).
7. Strausfeld, N. J., Strausfeld, C., Stowe, S., Rowell, D. & Loesel, R. The organization and evolutionary implications of neuropils and their neurons in the brain of the onychophoran *Euperipatoides rowelli. Arthropod Struct. Dev.* **35**, 169–196 (2006).
8. Mayer, G., Whitington, P. M., Sunnucks, P. & Pflüger, H.-J. A revision of brain composition in Onychophora (velvet worms) suggests that the tritocerebrum evolved in arthropods. *BMC Evol. Biol.* **10**, 255 (2010).
9. Whitington, P. M. & Mayer, G. The origins of the arthropod nervous system: Insights from the Onychophora. *Arthropod Struct. Dev.* **40**, 193–209 (2011).
10. Mayer, G. & Harzsch, S. Immunolocalization of serotonin in Onychophora argues against segmental ganglia being an ancestral feature of arthropods. *BMC Evol. Biol.* **7**, 118 (2007).
11. Mayer, G. & Koch, M. Ultrastructure and fate of the nephridial anlagen in the antennal segment of *Epiperipatus biolleyi* (Onychophora, Peripatidae)—evidence for the onychophoran antennae being modified legs. *Arthropod Struct. Dev.* **34**, 471–480 (2005).
12. Hirth, F. *et al.* An urbilaterian origin of the tripartite brain: developmental genetic insights from *Drosophila. Development* **130**, 2365–2373 (2003).
13. Eriksson, B., Samadi, L. & Schmid, A. The expression pattern of the genes *engrailed, pax6, otd* and *six3* with special respect to head and eye development in *Euperipatoides kanangrensis* Reid 1996 (Onychophora: Peripatopsidae). *Dev. Genes Evol.* **223**, 237–246 (2013).
14. Sombke, A., Harzsch, S. & Hansson, B. S. Organization of deutocerebral neuropils and olfactory behavior in the centipede *Scutigera coleoptrata* (Linnaeus, 1758) (Myriapoda: Chilopoda). *Chem. Senses* **36**, 43–61 (2011).

BRIEF COMMUNICATIONS ARISING

Cong *et al.* reply

In the accompanying Comment Mayer *et al.*[1] dispute ancestral affinities of *Lyrarapax unguispinus* and onychophoran brains. Here, we contest their claim that the evolutionary scenario described in Cong *et al.*[2] is unviable.

Mayer *et al.*[1] suggest that the frontal appendage ganglion in Extended Data Fig. 2e of ref. 2 is a 'sectioning artefact'. To demonstrate the frontal appendage ganglion, we present three consecutive silver-stained sections (Fig. 1a–c) of specimens from the data set[3] that Mayer *et al.*[1] accept as 'properly sectioned'. Figure 1c demonstrates the frontal appendage ganglion as being cytologically distinct from the nerve bundles entering it. Figure 1d shows axons entering it frontally and extending from it caudally. To illustrate that the frontal appendage ganglia do not exist, Mayer *et al.*[1] offer their Fig. 1a, which although labelled with anti-α-tubulin lacks the resolution to distinguish neuropil from axons. Their Figs 1b and 2a–d, claimed as 'full confocal projections', omit from the brain dorsal neuropils, optic tracts with optic neuropils, mushroom bodies (the brain's most prominent synaptic neuropil[3]), and lateral protocerebral regions. Their confocal images demonstrate that antisera useful for revealing axonal tracts, or selectively resolving peptidergic systems, fail to show neuropils revealed by silver or osmium-ethyl gallate staining[2,3].

To support their claim[1] that no neurons exist anterior to the protocerebrum, they offer Fig. 1c, in which sparse neuronal perikarya are stained presumably by dye leakage. Frontal appendages ('antennae' in Fig. 2b of ref. 1) show perikarya in front of the label 'at'. Developmental studies[4,5] also describe neurons and neuropil extending into the frontal appendage base.

Mayer *et al.*[1] state that if frontal appendage ganglia did exist they would be situated within the protocerebrum. Their statement that 'antennae' are protocerebral conflicts with a previous report[6] assigning to them a separate segment, thereby setting a precedent for introducing the term 'pre-protocerebral'. Mayer *et al.*[1] add confusion by stating that appendages, including the 'antennae', do not relate to segmental ganglia, citing the onychophoran central nervous system as asegmental[7]. If they subsequently claim[8] the protocerebrum as a brain segment, then appendages supplying it relate to that segment. The same applies to the onychophoran jaws, the nerves of which arise from the deutocerebrum[8] and whose claw-like shape[9] reveals an appendicular development and ancestry[10,11]. Identically formed claws define walking legs of Cambrian stem-group onychophorans[9]. However, onychophoran frontal appendages conspicuously lack claws, along with certain genes expressed in other appendages[12], suggesting that frontal appendages are either derived or that they antecede the evolution of other appendages.

Citing a pre-ocular location of chilopod antennae[13], Mayer *et al.*[1] argue that eye position is inadequate for resolving segmental identities. However, segmental affiliation relates not to an appendage's final post-developmental location on the head but to the brain segment supplying its axons. Chilopod antennae are supplied from the deutocerebrum. In Onychophora, frontal appendage axons supply paired centres anterior to the optic nerves, which define the protocerebrum. Hence frontal appendage centres are pre-protocerebral and accord with a frontal 'antennal' segment[6].

Figure 1 | *Euperipatoides rowelli* frontal appendage ganglia. Hemi-brains; anterior is up. **a–c**, Three consecutive silver-stained sections (ventral to dorsal) showing frontal appendage nerve bundles (fan) merging with the frontal appendage ganglion (frg) anterior to the eye (ey) and optic tract (opt), which supplies nested optic neuropils (on2) flanking the medial protocerebrum's (mpr) central complex (cc). The frontal appendage ganglion is confluent with the medial protocerebrum (border indicated with an arrow in **b**). Insets in **c** distinguish axons (fan) and neuropil (frg). **d**, Osmium-ethyl gallate-stained brain showing axons from the fan entering the frontal appendage ganglion and a broad axon fascicle (bracket) leaving it to flank the lateral protocerebrum (lpr) before entering the lateral nerve cord (lat nc), which corresponds to one of the paired descending tracts of *L. unguispinus* (dt in Fig. 2g of Cong *et al.*[2]). Bundled neurites (mb) supply the mushroom bodies, which originate ventral to this level. A small volume of mb neuropil is visible in panel **a**. Scale bars, 100 μm.

Peiyun Cong[1], **Xiaoya Ma**[1,2], **Xianguang Hou**[1], **Gregory D. Edgecombe**[2] **& Nicholas J. Strausfeld**[3,4]

[1]Yunnan Key Laboratory for Palaeobiology, Yunnan University, Kunming 650091, China.
email: xghou@ynu.edu.cn
[2]Department of Earth Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
[3]Department of Neuroscience, University of Arizona, Tucson, Arizona 85721, USA.
[4]Center for Insect Science, University of Arizona, Tucson, Arizona 85721, USA.
email: flybrain@neurobio.arizona.edu

1. Mayer, G., Martin, C., Oliveira, I. S., Franke, F. A. & Gross, V. Latest anomalocaridid affinities challenged. *Nature* **516**, http://dx.doi.org/10.1038/nature13860 (2014).

BRIEF COMMUNICATIONS ARISING

2. Cong, P., Ma, X., Hou, X., Edgecombe, G. D. & Strausfeld, N. J. Brain structure resolves the segmental affinity of anomalocaridid appendages. *Nature* **513,** 538–542 (2014).
3. Strausfeld, N. J., Strausfeld, C. M., Stowe, S., Rowell, D. & Loesel, R. The organization and evolutionary implications of neuropils and their neurons in the brain of the onychophoran *Euperipatoides rowelli. Arthropod Struct. Dev.* **35,** 169–196 (2006).
4. Eriksson, B. J. & Budd, G. E. Onychophoran cephalic nerves and their bearing on our understanding of head segmentation and stem-group evolution of Arthropoda. *Arthropod Struct. Dev.* **29,** 197–209 (2000).
5. Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanangrensis* with particular reference to the central nervous system. *J. Morphol.* **255,** 1–23 (2003).
6. Treffkorn, S. & Mayer, G. Expression of the decapentaplegic ortholog in embryos of the onychophoran *Euperipatoides rowelli. Gene Expr. Patterns* **13,** 384–394 (2013).
7. Mayer, G. & Harzsch, S. Immunolocalization of serotonin in Onychophora argues against segmental ganglia being an ancestral feature of arthropods. *BMC Evol. Biol.* **7,** 118 (2007).
8. Mayer, G., Whitington, P. M., Sunnucks, P. & Pfluger, H. J. A revision of brain composition in Onychophora (velvet worms) suggests that the tritocerebrum evolved in arthropods. *BMC Evol. Biol.* **10,** 255 (2010).
9. Smith, M. R. & Ortega-Hernández, J. *Hallucigenia*'s onychophoran-like claws and the case for Tactopoda. *Nature* **514,** 363–366 (2014).
10. Kennel von, J. Entwicklungsgeschichte von *Peripatus edwardsii* Blanch. und *Peripatus torquatus* n. sp. *Arch. Zool.-Anat. Inst. Würzburg* **8,** 1e93 (1888).
11. Sedgwick, A. *Peripatus* (Macmillan, 1922).
12. Oliveira, M. B. *et al.* Expression of arthropod distal limb-patterning genes in the onychophoran *Euperipatoides kanangrensis. Dev. Genes Evol.* **224,** 87–96 (2014).
13. Sombke, A., Harzsch, S. & Hansson, B. S. Organization of deutocerebral neuropils and olfactory behavior in the centipede *Scutigera coleoptrata* (Linnaeus, 1758) (Myriapoda: Chilopoda). *Chem. Senses* **36,** 43–61 (2011).